SIU | SOCIAL INVESTMENT UNIT

Social Housing Technical Report

June 2017

Measuring the fiscal impact of social housing services

New Zealand Government

SIU-2017-0089

Liability

While all care and diligence has been used in processing, analysing and extracting data and information in this publication, the Social Investment Unit gives no warranty it is error free and will not be liable for any loss or damage suffered by the use directly, or indirectly, of the information in this publication.

Citation

Social Investment Unit (2017). *Measuring the fiscal impact of social housing services – Technical report*, Wellington, New Zealand.

ISBN 978-0-473-39233-8 (online)

Published in June 2017 by

Social Investment Unit
Wellington, New Zealand

Contact

Social Investment Unit: info@siu.govt.nz
www.siu.govt.nz

Acknowledgements

The Social Investment Unit (SIU) would like to thank the following:

- Our joint venture partners Insights MSD (iMSD) and the Ministry of Social Development's (MSD's) Social Housing Policy team for their expertise in social housing.
- Statistics NZ for their consistency and generosity in supporting our numerous requests.
- All individuals who reviewed the draft technical report. The time and dedication involved is sincerely appreciated. All revision and feedback was invaluable and has resulted in this enhanced final version.

Disclaimer

The results in this report are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI) managed by Statistics NZ. The opinions, findings, recommendations and conclusions expressed in this report are those of the author(s), not Statistics NZ or other government agencies.

Access to the anonymised data used in this study was provided by Statistics NZ in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business or organisation. The results in this report have been made confidential to protect these groups from identification.

Careful consideration has been given to the privacy, security and confidentiality issues associated with using administrative and survey data in the IDI. Further details can be found in the privacy impact assessment for the IDI available from www.stats.govt.nz.

The results are based in part on tax data supplied by Inland Revenue to Statistics NZ under the Tax Administration Act 1994. This tax data must only be used for statistical purposes. No individual information may be published or disclosed in any other form, nor provided to Inland Revenue for administrative or regulatory purposes.

Any person who has had access to the unit-record data has certified that they have been shown and have read and understood section 81 of the Tax Administration Act 1994, which relates to secrecy. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes and is not related to the data's ability to support Inland Revenue's core operational requirements.

Executive summary

The Social Investment Unit (SIU) has completed its first test case of person-centred analysis to advance government's understanding of what is required to take a social investment approach.

Using data in the Statistics New Zealand Integrated Data Infrastructure (IDI), the test case estimated the fiscal impact of providing social housing support.

Building block to future social investment

Social investment is about directing and prioritising government activity towards measured improvements in people's lives. It requires:

- Establishing a data foundation
- Understanding the population and establishing what success will look like
- Identifying evidence-based interventions from the literature or previous experience
- Implementing appropriate programmes
- Monitoring and evaluating the impact of programmes and feeding this back into future decisions.

This test case related to *establishing a data foundation*, and *evaluating the impact of programmes*. Related work on segmentation of social housing recipients, to be published later in the year, looks at *understanding the population*. Analytic work such as this is necessary but not sufficient for a social investment approach. This test case should be seen as a building block to future social investment programmes, and not limited to social housing.

Objectives of this report

The SIU aimed to understand whether it was possible to calculate a fiscal return on investment (ROI) for a given intervention within the social sector. This test case looked into social housing – provided by Housing New Zealand – as a case study to test this.

The objectives were to:

- Understand whether it was possible to calculate a fiscal ROI for a given investment within the social sector
- If so, develop a reusable methodology that allows the analysis and dataset to be re-used
- Understand the methodology's limitations.

Closer to understanding fiscal ROI

Of the Crown's $50 billion annual spend on Social Development, Health, Education, Police, Justice Courts, Corrections and ACC, an estimated $33 billion could be associated with individual New Zealanders. However, it was not possible to connect spend of $17 billion with individuals; much of this spend was system-wide, while some of it simply requires further work to estimate the spend relationship to individuals.

Two groups were compared for this test case: social housing applicants who were successful and those who were unsuccessful. Statistical methods were used to make valid comparisons between these two groups.

The focus was on demonstrating the method rather than findings for concrete action.

Tentative conclusions

While it is not yet possible to estimate a comprehensive fiscal ROI for social housing due to limitations in the data, analysis indicated:

- People in social housing have **25% *less* spend from Corrections** ($13 million), than those not in social housing.
- Children in social housing have **6% *more* education spend** ($16 million) than those not in social housing.
- People in social housing received **3.6% *more* main benefits** ($31 million) – i.e. employment support, sole parent and sickness benefits – than those not in social housing.

Reusable methodology and tools

Developing estimation tools to allocate social spend to individuals was a key step towards creating reusable methodologies. This included producing and publishing a computer code repository – the **Social Investment Analytical Layer (SIAL)** – to reshape some of the IDI data. The SIAL is available on the code-sharing website GitHub.com.

The SIAL gives a head start for future analysis needing to allocate 'events' and their costs to individuals.

A companion tool – the **Social Investment Measurement Map (SIMM)** – lists person-centred outcomes that can be measured in the SIAL, and aims to help future analysts unfamiliar with the IDI data to gain a better appreciation of what is available.

Work on a general approach to data preparation in the IDI was started, using the SIAL as a key step. This will be published later in 2017, allowing analysts to structure data preparation in a few simple steps (such as defining cohorts of interest, intervention type, explanatory variables and risk factors, and outcomes of interest in addition to fiscal costs), and quickly move on to analysis.

Using the fiscal cost data made available by the SIAL, the project also confirmed the statistical method of 'inverse probability of treatment weighting' can successfully be part of an analytical approach to estimating fiscal impacts of a government intervention.

The SIAL and SIMM are available on the SIU's website www.siu.govt.nz.

Limitations better understood

There were **important fiscal costs that were not able to be allocated to individuals** because the data was not available in the IDI, e.g. data on primary health services (including visits to general practitioners) and many other services.

In the case of social housing, it is expected these absences made a material difference to conclusions about the total fiscal impact. Work is underway to improve data coverage so the IDI can give a fuller picture in future.

Even when data was available and could be allocated to individuals, **relatively crude assumptions about costs** had to be made for many government services. Often this meant allocating average costs to a wide group of individuals simply on the basis of length of engagement with a service. These costings can – and will – be improved over time through better estimation methods and if better costing data can be directly integrated into the IDI.

The focus on fiscal ROI is insufficient for many purposes of comparing investment options. This is not a method that aims to replace traditional approaches, such as cost benefit analysis of net change and cost-effectiveness of quality-adjusted life years.

Our recommendations include further work on applying similar methods to those in this project to a wider range of non-fiscal outcomes.

Recommendations for future work

Future work by the SIU will address the first three recommendations from this project – outcomes, well-being and segmentation – as their general applicability gives them priority.

- *Examine and monitor social outcomes in more detail*: this will paint a more accurate picture of the effective impacts of social housing (and future social policy investigations). Focus should be on exploring social and economic returns, such as educational attainment, employment rates, child abuse rate, etc.

- *Define and use a well-being framework:* outcomes could be coupled to a 'human-centric' framework that would produce an agreed measure of social well-being. Social, economic and cultural ROI measures would complement fiscal insights.

- *Discriminate results by profile*: address the difficulty in pinpointing who would benefit most from social housing (and future investigations). A segmentation exercise would identify profiles of people. Once profiles are identified, the method could be reproduced rapidly on each of the segments to measure the returns and behaviours exhibited by each.

The remaining recommendations are specifically related to social housing. They may be addressed if (and when) future analysis is undertaken on social investment for social housing.

- *Build a predictive model*: estimate the monitored cost (either total or per item), regarding the detailed characteristics of the household. This would be difficult but the result would be more detailed and useful.

- *Take length of tenure into account*: the training set (showing monitored costs with regards to household characteristics only), would likely show too large a variation to allow an accurate model to be built.

- *Take changes in household composition into account*: monitor at an 'individual' level rather than 'household' level.

- *Consider a long-term (even lifetime) forecast window* for monitoring both comparison and treatment groups – this would allow the effect of observed fiscal impacts to be measured, e.g. higher education costs.

Contents

1 Introduction/context

1.1 Social investment

Social investment is about improving the lives of New Zealanders. It is called *investment*, not spending, because it is about investing resources upfront to enable people in need to thrive over the long-term. It puts the needs of people who rely on public services at the centre of decisions on planning, programming and resourcing by:

- Using information and technology to better understand the needs of people and the services they are currently receiving
- Systematically measuring the effectiveness of interventions to understand what works for whom and at what cost
- Understanding the fiscal implications of better outcomes and helping to manage the long-term costs to government
- Funding to the most effective services irrespective of whether they are provided by government or Non-Governmental Organisations (NGOs)
- Getting better outcomes for targeted populations, particularly the most vulnerable
- Making a positive impact on the lives of those New Zealanders most at risk of poor outcomes – children, young people and adults.

Insights gained can then be fed back into the decision-making process to make better-informed, evidence-based policies and social services. Investing in the services we know work and managing the long-term costs to government will improve the lives of New Zealanders.

Social investment signals a shift in the Government's approach to social spending.

1.2 The Social Investment Unit

The Social Investment Unit (SIU) was established in April 2016 as an independent cross-agency unit responsible for implementing the Government's social investment approach. The SIU supports the aim of putting analytics at the core of decisions on government spending by applying rigorous and robust data-driven methods of evaluating the effectiveness of social policy outcomes. The SIU:

- Works with agencies to deliver the tools and infrastructure required to support a social investment approach
- Provides independent cross-sector advice.

1.3 The Social Housing Test Case

1.3.1 Why social housing?

Social housing services were chosen as the focus of the SIU's first test because it is at the centre of wider, cross-agency work, including:

- Government budget decisions (Treasury)
- Social housing purchasing decisions (MSD) and Housing New Zealand (HNZ)

- Social housing needs assessments (MSD)
- Social housing valuation (MSD)
- Social investment advice (SIU)
- Social housing scenario modelling work (SIU and MSD).

In addition to the goals listed above, the research aims to generate insights into the impact of social housing on broader social sector spending, particularly on services which can have a positive impact on people's lives.

1.3.2 Social housing context

Government and NGOs provide many housing support services. These include:

- The Accommodation Supplement (AS) to assist with housing costs in the private market
- Temporary Additional Support (TAS) to assist with significant financial hardship
- Emergency housing
- Social housing.

Current social housing provision involves:

- Access to a physical house, whether provided by HNZ or a Community Housing Provider (CHP)
- Provision of the Income-Related Rent Subsidy (IRRS), where clients pay an Income-Related Rent (IRR) below the level seen in the private rental market.

Social housing is primarily targeted at people with very high housing needs unable to be met by the private market. Social housing addresses housing affordability, as well as helping those unable to access private rental housing for a range of other reasons such as discrimination or health issues.

Clients are scored against the Social Allocation System (SAS), the five main categories of which relate to:

- Affordability
- Adequacy
- Suitability
- Accessibility
- Sustainability.

MSD is developing an investment approach to social housing which measures the forward liability associated with the SAS. The liability acts as a proxy for assessing people's risk of long-term social housing dependency and provides a tool to assist management in working with clients.

The analysis described in this technical report is another component of the social investment approach to social housing.

1.3.3 Social Investment approach to social housing

Social housing is provided to those with severe and urgent housing needs. It can have a protective effect by supporting people in crisis who need accommodation immediately, and who may also face a range of social problems. Social housing is intended to help:

- Women leave violent relationships (family violence, maltreatment outcomes)
- Prisoners reintegrate and lead crime-free lives (reoffending outcomes)
- People manage their health issues (mental health outcomes).

In addition, social housing can impact social outcomes in positive ways, providing a return on the government's investment by:

- Freeing up a household's resources via the immediate impact of the subsidy on housing costs and the longer-term impact of improved employment outcomes
- Improving children's long-term health, education and social outcomes via:
    - Flow-on effects of positive change in a household's financial position, (could impact on parenting)
    - The impact of positive changes in housing conditions, e.g. better housing quality and safety, less household 'crowding' and improved neighbourhood characteristics
    - Residential movement and transience, e.g. the quality of the household's social support networks and a child and family's connection with health and education services.

These positive social impacts create fiscal costs and savings, which can be considered alongside the cost of social housing provision itself (rent subsidies, capital costs, tenancy management and other administrative overheads).

These costs and savings clearly highlight the value in deriving a wider, cross-agency ROI figure. In particular a fiscal ROI, focusing on differences in spends by and revenues to government agencies towards social votes provides an estimation of these social impacts.

1.3.4 Business value

This is the first time the costs and benefits for those receiving social housing support have been quantified across agencies.

The analysis will assist government agencies quantify the impact of investment in social housing on other areas of social sector spending. This will help:

- Understand where the costs and benefits of living in a social house accrue across various government agencies, as far as can be measured in the data
- Target and prioritise decisions – by understanding the impact on outcomes for different kinds of groups, the analysis can inform which applicants would benefit most from social housing, and pinpoint:
    - What information to collect through the SAS – the needs assessment framework used by HNZ

- Which groups to prioritise on the basis of that assessment.

The analysis in this report did not inform the SAS review performed in November 2016 by MSD.

As stated, the general purpose of the analysis was to understand the possibility of a fiscal return to government across social sector agencies for people who receive social housing support. The analysis measured the Average Treatment Effect on the Treated group (ATT) in terms of fiscal impact – related to the housing services that were provided by HNZ only between 2005-2006, not all social housing.

The returns calculated in this report are fiscal-only and are based on government administrative data contained in Statistics NZ's Integrated Data Infrastructure (IDI). The IDI has been chosen as the source of data as it is currently the best integrated data source available for New Zealand's population across time. Although the focus of the present work was on fiscal ROI only, it will be important to develop social, economic and (if appropriate) cultural ROI measures to complement the fiscal insight. These developments were not within the scope of this work. SIU intends to explore social and economic returns in future test cases.

1.3.5 Relationship with existing work

Other components of MSD's social housing social investment approach will include a valuation of social housing households (similar to that done for the welfare system), and the implementation of key performance measures.
The social housing valuation estimates the lifetime social housing costs for those in social housing and the notional lifetime social housing cost of people on the social housing register. Liability is estimated in terms of the cost of social housing itself, including AS and TAS for people included in the valuation.

Preliminary work was done by Insights MSD[1] (iMSD) on a scenario analysis tool that would enable measurement of changes to efficiency and effectiveness for a range of policy choices. Currently the framework has been set up and tested using synthetic data. There is potential for this to be progressed into an interactive tool for social housing.

The work of SIU and MSD's social investment component complements the valuation work in the following ways:

- The social investment approach uses data from all agencies available in the IDI – the valuation uses administrative data from MSD
- The ROI is intended to be a robust and unbiased assessment of the effectiveness and impact of social housing services that have been provided. The valuation is a forward liability model
- MSD's social housing framework is designed to provide housing for the right person, in the right place, for the right duration. The valuation work focuses primarily on estimating lifetime costs for people who receive social housing assistance, while the SIU/iMSD work focuses on an individual's life trajectory to target the right people to support with social housing by measuring impact on spend across their lives, i.e. education, health, employment etc.

[1] Insights MSD (iMSD) is the Ministry's center of expertise for research and analytics. The purpose of iMSD is to maximise the value of the Ministry's data and analytics capability to drive better outcomes for its clients.

This work creates a fuller picture for decision-makers to use.

1.4 How the SIU works

SIU intends to support agencies in their embracing of a social investment approach by deriving and sharing robust analytical methods. It is therefore essential the quality of its work is verified by independent experts.

The initial analytical work was conducted by the Evidence and Insights team between July and October 2016. Todd Nicholson Consulting was engaged over this period to provide technical advice and support.

Preliminary results where shared early with several subject matter experts, policy analysts and data analysts from MSD and the Treasury, seeking feedback and comments. An independent review was conducted at the same time by Sapere Research. Valuable feedback was received from these parties, a number of issues were addressed and clarifications were made. When it was not possible to do address these earlier, limitations of the current approach have been clearly highlighted within this report.

This review phase was enhanced by sharing the code and methods developed.

1.5 Outline of the report

The Social Housing Test Case is designed to evaluate whether it is possible to measure the effectiveness of a given intervention in terms of cross-sector fiscal spend. To do this, a *propensity score-matching strategy* was used to create a counterfactual group to be compared to the group of people who received social housing assistance within the timeframe examined.

Section 2 deals with this methodology – the approach and the details. This approach to evaluating the effect of a treatment received refers to an ATT analysis. The early sections relate to this evaluation. Later sections describe how the cohort at the focus of the test case analysis was built and also provides related descriptive statistics.

Section 3 reports on the propensity score model built to estimate the probability of an individual receiving a given treatment (social housing support). The section focuses on the model, its training and performance. Also detailed is the application of this method to derive the treatment and comparison groups.

Section 4 focuses on the necessary steps, and the limitations, of calculating a single fiscal ROI figure, while Section 5 reports on the computed results.

Section 6 details the questions generated and directions identified for future work.

Section 7 contains recommendations.

2 Methodology

2.1 Introduction to the methodology

The aim of this test case was to identify the fiscal impacts of receiving social housing services, as measured by the differences in government spends on two cohorts of households:

- Treatment group – those who received social housing support
- Counterfactual group – those who did not receive social housing support.

In the context of this analysis, the definition of social housing support was restricted to the provision of a social house by HNZ, by opposition to the payment of AS or to the provision of housing by alternative entities (e.g. community or council housing).

The crucial point of such analyses is to ensure the two groups being compared differ only by their receiving social housing support or not. It is important the characteristics of the two groups are as similar as they can be.

If a Randomised Control Trial (RCT) can be conducted then this can be ensured. However, in many contexts (including housing) it is not appropriate to conduct such an experiment and randomly assign an intervention. Further, when analysing past interventions, it is not possible to construct an RCT.

In this test case, it was only possible to rely on the observation and analysis of people (households) that have already received or been denied social housing support. Because the two groups of households present some inherent differences in their features, a mathematical treatment needed to be applied before the comparison was made to create comparable or *balanced* groups.

Propensity score methods are used to derive comparable groups from observed, non-random data [see Austin *et al*. (2011)]. These methods create comparison groups by identifying matches to the individuals in the treatment group based on their propensity score, that is, their estimated probability of receiving the treatment.

Several different matching methods exist to do this, including:

- Calliper matching
- Radius matching
- Nearest neighbour matching
- Sub classification.

Through testing it was found the results produced by the different matching methods were very similar and the choice of the matching algorithm depended on what an individual organisation tended to use. Treasury has published several reports using calliper matching, while unpublished work from MSD used nearest neighbour matching with replacement.

The test case used *Inverse Probability of Treatment Weighting* (IPTW) [see Austin *et al*. (2015)]. One advantage of IPTW is there is no loss of subjects, as occurs with matching algorithms. Similar to the idea of using survey weights, IPTW adjusts for differences in probabilities by attributing a *weight factor* to individuals in the cohort. It uses the predicted probability (propensity score) of obtaining the actual treatment a subject received.

Note a known limitation of propensity score matching methods is they condition on observable differences only – it is assumed this also controls for unobserved differences. If this assumption holds, then the resulting estimates are unbiased.

2.2 Estimating the Average Treatment Effect on the Treated (ATT)

In the context of ATT, individuals in the treatment group are given a weighting of 1. Subjects in the counterfactual group must be given weights so the weighted distribution of propensity score matches the one of the treated group. The weighting coefficient is set equal to their predicted probability (of being housed), divided by 1, minus this probability. This weighting strategy has the desired upward effect on individuals with high propensity scores (which are under-represented in the comparison group), and the downward effect on low propensity scores.

In a more formal way, denoting X_i the vector of characteristics (features) on an individual i, Z_i their class (1 for housed, 0 for not housed), and Y_i the measured outcome (in this case, the fiscal benefit or cost measured on the outcome window). The ATT, estimating the difference in expected outcome averaged over all profiles, is given by:

$$ATT = E[Y|X, Z = 1] - E[Y|X, Z = 0]$$

If a total population is considered of $n=n_T+n_C$ individuals (n_T being the number of people housed and n_C the number of people not housed), the formula to estimate the ATT is:

$$ATT = \frac{1}{n_T}\left(\sum_{i=1}^{n_T} Z_i.Y_i - \sum_{i=1}^{n_C} w_i.(1 - Z_i).Y_i\right)$$

Where:

Z_i denotes the class the individual belongs to, set to 1 for the treatment group (housed) and 0 for the comparison group (not housed) .
Y_i denotes the effect or outcome monitored – in this case the measure of fiscal benefit (hence negative if it is a cost), measured on the outcome window.
w_i denotes the weight (calculated as the inverse probability of treatment).

It follows that if the estimated probability of receiving the treatment (being in social housing) is noted p_i and if the treated population is indexed with index $i=1,\ldots n_T$ and the comparison population with index $j=1,\ldots, n_C$, then the formula above becomes:

$$ATT = \frac{1}{n_T}\sum_{i=1}^{n_T} Y_i - \frac{1}{n_T}\sum_{j=1}^{n_C} \frac{p_j}{1 - p_j}.Y_j$$

In this equation, $p_j/(1-p_j)$ is the effective inverse probability weight. It is assumed the sum of weights is approximately equal to the number of people in the comparison group. If that is not the case, a scaling factor equal to $n_T/\sum_j w_j$ can be applied to correct for the difference in the two population sizes.

2.3 Creating the cohort and groups

To apply the ATT method, the first requirement is to determine a population of interest (or cohort), as well as a timeframe over which the population will be monitored, both before and after their receiving the treatment. Specifically, the following questions were asked when constructing the population:

- How to identify and link people across different agencies?
- Who is included in the chosen group of people to model?
- How long is this group followed into the future to calculate ROI?
- How far back is this group looked at in terms of other factors that might have influenced their housing application?

2.3.1 Source of data

The IDI was used to access cost information from various agencies (Inland Revenue (IR), Accident Compensation Corporation (ACC), Ministry of Health (MoH), MSD, Child, Youth and Family (CYF), Ministry of Education (MoE) and Department of Corrections (COR)) for individuals.

An individual will have a series of interactions with various government agencies throughout their lifetime, referred to as events. Tables coming from the government agencies listed above capture these events and report information on start and end dates, type of events and costs. Most events tables have good coverage between 2001 and 2014. Some events tables have longer coverage, some less.

Each government agency uses a different set of unique identifiers to identify people and record events within their system. Using different sets of identification numbers (IDs) makes it difficult to create a set of cross-sector events for a single person. Events for a person are linked in the IDI using what is referred to as the 'spine'. The spine aims to identify each person once by using IR numbers, the birth register and immigration data.

However, it is possible to find events tied to a person who is not linked to the spine – this is most likely caused by linkage error. To ensure only genuine people were included in the test case cohort, any applications with individuals who were not attached to the spine were filtered out.

2.3.2 Unit of analysis and time period of interest

A unit of analysis needed to be agreed before doing the analytical work. The test case looks at the impact of social housing for the *household* rather than the individuals within it. The household was defined, for the purposes of this report, as those who feature on the housing application form. This is because the social housing mechanism looks at those who are mentioned on the applications.

Household composition changes over time. This fact was not accounted for over the short timeframe of the analysis.

On the basis of data availability, it was decided the cohort of interest for this analysis would comprise all those households *who applied* for *social housing between 1 January 2005 and 31 December 2006*, and whose application was cleared from the social housing register

within two years of application date. Exit from the register occurred either because they received a social house from HNZ, or because their application was declined, or because they voluntarily cancelled their application, e.g. if they found an alternative housing arrangement such as community or council housing).

The rationale for picking this time period was it would provide a four-year window of past data to define variables that characterise the outcome of the application. Additionally, by putting the two-year restriction on application exit date, an adequate *follow-up period of at least six years* would be available to measure the outcomes from the intervention.

This additional condition does not significantly affect the number of applications under consideration, as roughly 95% of the applications submitted in 2005/06 were exited within two years of application date.

The cohort of interest includes both the applications that eventually received social housing support and those that were rejected. The housed applicants became the treatment group and the rejected applicants became the counterfactual group.

For the *treated* group, the outcomes were measured starting from the date of *approval* for social housing, whereas for the *counterfactual* group, the outcomes were measured from the *date following the application*.

This introduced a potential for bias, since the outcomes were measured from two different time points for each group. To quantify the potential for bias, the average time for a housed application between application date and affective housing was inspected and found to be only around 82 days (the median value is 39 days). Hence, it can safely be assumed the conditions of these households remained relatively unchanged, on average, from the date of application to the date of housing.

2.3.3 Summary of rules for building the cohort

The detailed business rules for defining the cohort were:

- All HNZ applications submitted between *1 January 2005 and 31 December 2006*, and which had an *exit date of at least within two years of application,* were in scope. Applications for transfer were excluded.
- Of these applications, all individuals who were part of the application needed to be linked to the IDI. If any individual included under an application could not be linked to the IDI, the application was filtered out from the cohort.
- In cases where an individual was part of multiple applications with differing exit status, all applications to which that individual was linked were discarded. This ensured everyone in the cohort had only a single exit status in case of multiple applications.
- In cases where an individual was part of multiple applications with the same exit status, the earliest application by application date was retained. In cases where multiple applications were submitted on the same day for the same individual, the application with the larger ID number was retained.
- If the resulting set of applications created a single individual being part of multiple applications, such applications were removed to ensure the cohort included every individual only once.

This produced a group of approximately 22,000 applications, where approximately 11,000 received social housing support and the remaining 11,000 did not.

The sums in the tables below (see section 2.4.2) may not equal this total due to confidentiality rounding.

2.4 Characteristics

2.4.1 IDI data and the Social Investment Analytical Layer (SIAL)

Data in the IDI comes in various formats, reflecting the standards and formats used by the source agencies. Unfortunately this does not facilitate an automated treatment and preparation of data for analysis. To overcome this issue, data sets from various sources were standardised into a common PTCE (Person Time Cost Event) format.

Events tables generated by SIU in this common format and the methods used to construct them have been shared with the IDI community for future use. The tables, formatted to facilitate analytics studies, are referred to as the Social Investment Analytical Layer (SIAL).

Once the standardised tables were constructed, *standardised code* was applied to extract variables of interest. Household characteristic variables, as well as expert variables for HNZ applications, particular to this test case, were derived. Characteristics related to the primary applicant were also used. Descriptions of all the variables used can be found in Appendix A: Descriptions of variables used.

Detailed information (and an overview) about the SIAL can be found in Appendix B: The Social Investment Analytical Layer (SIAL).

2.4.2 Descriptive statistics

A selection of descriptive statistics follows, with the remainder contained in Appendix C: Cohort descriptive statistics. These are actual, non-weighted counts and relate to the primary applicant. Detailed counts may not add up to the total due to rounding – for confidentiality reasons.

The 'Region' variable was sourced from many different datasets. If a region could not be found it was placed in the 'unknown' category.

By 'Received social housing' and 'Did not receive social housing'

HNZ exit status	Count
'housed'	10,629
'other exit'	11,193

By age band

Age band	Housed	Other exit
0-23	1632	2106
24-29	1824	1815
30-34	1536	1593
35-40	1662	1689
41-49	1656	1737
50-64	1548	1515
65 and over	771	738

By prioritised ethnicity (primary applicant)

Prioritised ethnicity (primary applicant)	Housed	Other exit
Asian	486	594
European	3051	3951
Maori	4743	4497
MELAA	396	393
Pasifika	1911	1674
Other	48	90

***** Middle Eastern/ Latin American / African

By gender (primary applicant)

Gender (primary applicant)	Housed	Other exit
Male	3387	3498
Female	7245	7698

By current region code

Region code	Region	Housed	Other exit
1	Northland	459	708
2	Auckland	3837	3603
3	Waikato	846	957
4	Bay of Plenty	501	636
5	Gisborne	234	153
6	Hawkes Bay	543	690
7	Taranaki	237	339
8	Manawatu-Whanganui	507	594
9	Wellington	1569	1053
12	West Coast	72	114
13	Canterbury	1008	1329
14	Otago	351	315
15	Southland	135	129
16	Tasman	30	78
17	Nelson	84	192
18	Marlborough	78	153

Region code	Region	Housed	Other exit
98	Unknown	144	147

By income category

Income band	Not received social housing	Received social housing
No income	3147	2982
Up to $5,000	5100	5805
$5,001-10,000	1281	1335
$10,001-$20,000	960	924
$20,001-$30,000	126	120
$30,001-$40,000	18	27
$40,001-$50,000	0	3
$50,001+	3	0

By household size

Household size	Not received social housing	Received social housing
1	2862	3873
2	2697	3495
3	2136	2055
4	1443	939
5	819	450
>=6	675	378

2.5 Computing and monitoring costs

As the aim of this analysis was to test the possibility of devising a measure of fiscal ROI for government spending on social policy, it was first necessary to understand and list the cost information available in the IDI.

The IDI contains many fiscal costs and benefits and is a great source of data for monitoring social spends, but it is incomplete. The Social Vote accounts for around $50b (as at 2014) of Government spend. Within the IDI, approximately $33b can be attributed to social spend at an individual level (see Figure 1).

The agency lacking the largest amount of costs available in the IDI is MoH, with only 40% of spend being attributed to an individual. With a longer research timeframe, this could have been improved by using additional (external) MoH information to create derived costs.

Some data included direct costs associated to events. In other cases, these costs had to be derived.

Some systems contain slightly overlapping information. For example, MSD tier 1[2] is similar to what is coded BEN (benefit) in the IR tables. The main difference is that MSD data is based on *entitlement,* whereas IR data is based on what is actually *received*. Consequently, in place of MSD tier 1, the costs from IR have been taken and relabelled as MSD.

2.5.1 Derived cost/return information in the IDI

This section outlines the costs from the IDI output that were derived and the rationale and process for doing so.

All tables other than the ones listed in the remainder of this section (Hospitalisation, PRIMHD, Corrections and Education) below have 'cost by event' information available in the IDI, so there was no need to derive these costs. Derived costs were only used when there was no costs information readily available and when a sound method to derive them could be identified.

2.5.2 Hospitalisation outpatient events

Hospitalisation outpatient events do not have costs, but they do have purchase units. The cost per purchase unit was sourced from an external source and applied to the purchase units in the IDI.

An event can be a hospitalisation or a non-admitted event. Non-admitted events are either labelled Emergency Department attendance (ED), Outpatient event (OP) or Community event (CR).

The prices for hospitalisations were based on a WIES cost weight, multiplied by a medical-surgical price.

The non-admitted events were based on a nationally-contracted price related to the purchase unit.

Contracted price and unit med-surgical price were provided by MoH – District Health Board (DHB) monitoring and performance.

2.5.3 Mental health (PRIMHD) events

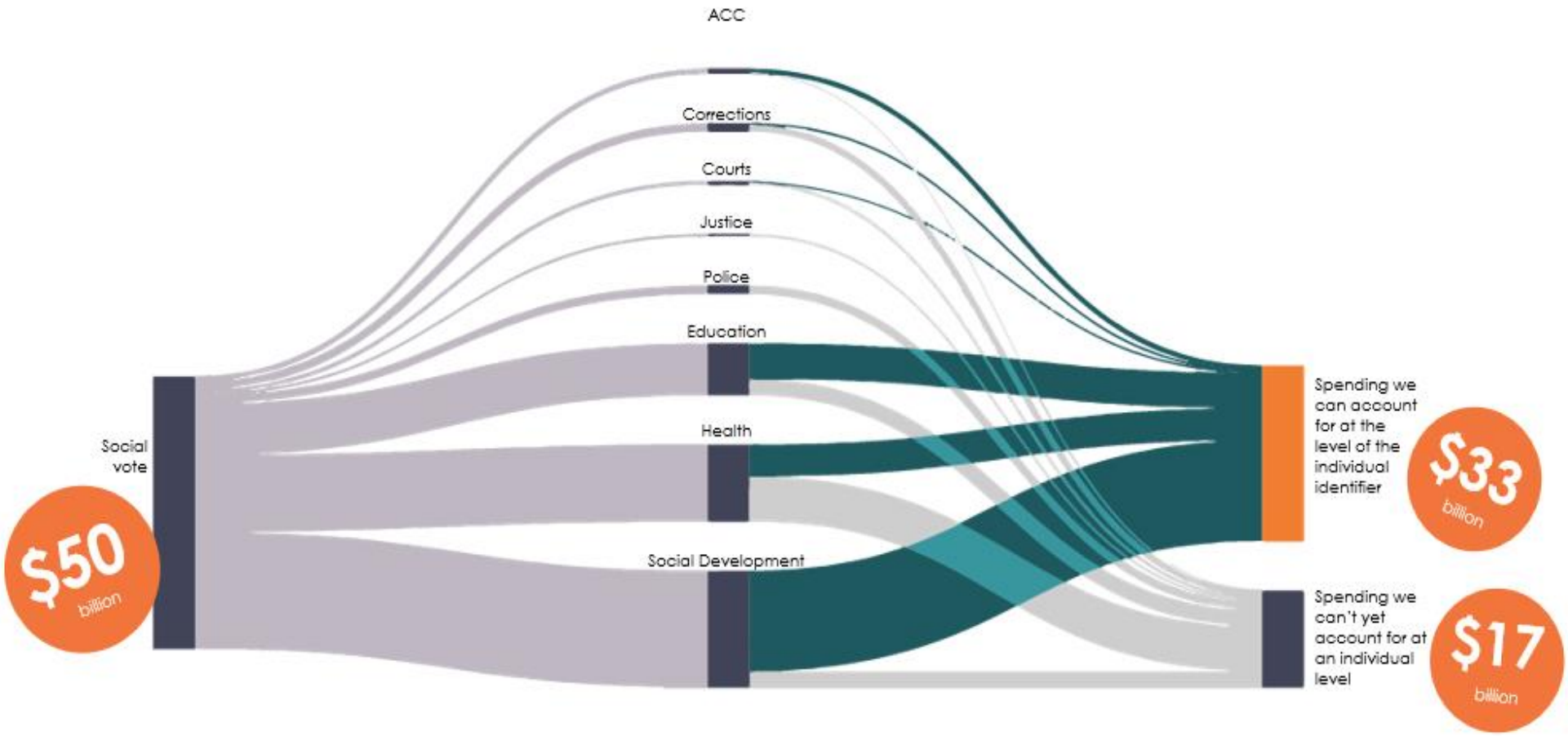
Mental health datasets have a yearly cost per person. There are three different event types each person can have in a year: bed nights, contacts and seclusion.

The total yearly cost per person was split between these three events using a regression model. Where a person did not have one of the three event types in a given year, their costs for that event are zero and the remainder of the cost has been split between the other two events.

At a total level (all mental health events for a person), these costs will be correct (assuming the IDI is correct), as they are not derived. When specific event types for mental health are used, the costs are derived.

[2] First tier benefits refers to main benefits such as Job Seeker Support (JSS) and Sole Parent Support (SPS) and others, that are expected to meet basic living costs

Figure 1: Amount of the social vote accounted for in SIU analytical layer in the IDI 2014[3]



[3] Note, these figures have been rounded for presentation purposes.

2.5.4 Corrections events

Corrections costs were supplied by offence type. These were added to the Corrections events in the IDI and a cost was derived.

2.5.5 Education

The data provided by MoE[4] was used to calculate a mean annual value of funding for salary, operational and direct property per student (for both state and state-integrated schools by school decile and year). These values were applied at the individual student level on a pro-rata basis dependent on length of enrolment at each school. Private school funding of these types was set to zero.

2.5.6 Monitored costs: summary

Table 1 summarises all cost items monitored over the six year follow-up period to compute the total fiscal social spend associated to households.

Table 1: Monitored costs per agency and subject area

Agency or Ministry	Subject area	Corresponding costs
ACC	CLM	ACC weekly compensation claims
ACC	INJ	ACC non earner medical costs
CYF	CNP	Care and Protection
CYF	YJU	Youth justice
COR	SR	Sentencing & Remands
IR	PPL	Paid Parental Leave
IR	STU	Student Allowance
MoE	ENR	Student Enrolment
MoE	B4S	Before School Check
MoH	GMS	General Medical Subsidy
MoH	NNP	National Non-admitted patient collection
MoH	PFH	Publicly funded hospitals
MoH	PHA	Pharmaceutical
MoH	PRI	Mental health (PRIMHD)
MoH	TES	Lab test
MSD	T1	Tier 1 benefits
MSD	T2	Tier 2 benefits*
MSD	T3	Tier 3 benefits

* Excluding Accommodation Support

2.6 Key lessons and future improvements

Understanding the business problem and converting it into a statistical problem was one of the most challenging aspects of this research. Understanding data quality issues and the business context around the variables is also crucial.

[4] http://www.educationcounts.govt.nz/statistics/schooling/resourcing/47696

Key lessons learnt during this process and recommendations are:

- Using the spine when constructing the population from IDI data is one method of ensuring IDs are appropriately linked and refer to an actual person or entity. Future options include using the Estimated Resident Population (ERP) constructed by Statistics NZ to identify who is in New Zealand. However, this appears to be a recent ERP, so it cannot be used in studies assessing past interventions.

- Defining the study population requires careful thought so it remains relevant to the business, yet it needs to be simple enough to model so initial results can be seen. The simplifying assumptions for social housing were to look at people housed within two years and to exclude transfers. The decision to look at those who were housed within two years was to guarantee *each household had six years of costs measured*. This ensured differences due to the length of time over which the costs were measured were not seen. The decision to exclude transfers kept the analysis and interpretation simpler. Looking at the first time a person is housed is still beneficial. However, because a cross-section of time was viewed, it could not be guaranteed all applicants had not been housed before.

- People can apply multiple times for social housing but it would require more complicated modelling to identify differences. The simple solution was to remove any duplicate applications. This meant only the differences social housing made the first time round, within the given timeframe, were looked at. Therefore, totals in the results section are for a particular part of the 2005/06 population. Applied to the whole population, the differences at a total level would differ.

- It is important to be wary of case-comparison pollution. This arises when a household applies for social housing multiple times and ends up unsuccessful in one application and successful in another. This problem was avoided by keeping the successful application (applicant placed in a house) and discarding any others.

- Feedback received questioning the original descriptive statistics greatly improved this section of analysis. However, the majority of people do not know the details of the social housing application process. Labelling and commenting before presenting descriptive statistics makes the work more digestible and allows other groups to carry on future work more easily.

- Consult frontline experts early in the project, not halfway through. Look for experts in business process and policy from the relevant agencies, in this case experts from HNZ, since the cohort was from a HNZ assessment conducted in 2005/06.

- Care needs to be taken when describing variable transformations and imputations. The first version of the report adequately described transformations and imputations in the appendices but the body of the report gave the erroneous impression all the missing regions were imputed as ‘Auckland’.

- The unit of analysis can make variable construction tricky. For example, the unit of analysis for social housing was the household, but the SIAL tables produce measures at the individual level. These were then aggregated to the household level. Some of the descriptive statistics would have been more comparable if they had been transformed into things like equivalised income. This limitation can be explored in future test cases.

3 Propensity score modelling

The previous sections describe how the two populations monitored and compared (treatment/housed and counterfactual/not-housed) were built. As detailed, a propensity score method was used to derive comparable groups from observed, non-random data. Such methods rely on a model giving the estimated probability of an individual (or more generally, a unit of analysis – a household, in this case), receiving a treatment based on a set of identified characteristics – the propensity score.

This section reports on the propensity model built for this purpose. The model estimates the probability of a household having lodged an application for a social house being granted one. The attributes considered as inputs (covariates) of this model are described below, as well as their pre-processing and transformation.

3.1 Variable transformation and pre-processing

3.1.1 Initial variable description

This section describes each of the variables considered in the model for predicting whether a social housing application results in a social housing placement. The variables are classified into three major categories:

- HNZ application-related
- Primary applicant characteristics
- Household-level characteristics.

Not all variables listed were used in deriving the probability scores for receiving social housing support. Instead, a subset of these variables was used on the basis of:

- Subject matter expertise
- Data quality
- Data transformations
- Automated feature selection.

Full descriptions of these variables can be found in Appendix A: Descriptions of variables used.

3.1.1.1 HNZ application variables

These variables are based on HNZ's evaluation of the social housing application and information provided by the applicant as part of the application:

- Accessibility score
- Adequacy score
- Affordability score
- Application main reason
- Bedroom count required
- Current region code
- Size of household
- Suitability score
- Sustainability score

- Total score
- Quarter and year of application.

3.1.1.2 Primary applicant characteristics

These variables relate to the characteristics of the primary applicant making the social housing application:

- 12-month wage to benefit ratio (log transformed)
- Age category (age split into bands – see Appendix G)
- Prioritised ethnicity
- Gender.

3.1.1.3 Household characteristics

These variables describe the characteristics of the household/family that has applied for a social house.

Data is consolidated from a variety of sources, including CYF, benefits data from MSD, health and medical services data from MoH, MoE, COR and claims and injuries from ACC:

- Yearly Household Income
- Accidents/Injuries – ACC Claims-related costs
- CYF Yearly Abuse Event-related costs
- YJU Yearly Incidents-related costs
- COR Sentencing and Remands costs
- MoE Student Interventions Count
- MoH Cancer Registration Events Count
- MoH Chronic Conditions Registration Events Count
- MoH Yearly General Medical Subsidy Claims Amount (two years before application)
- MoH Yearly Hospitalisation Costs (two years before application)
- MoH Yearly Lab Costs
- MSD Yearly Tier 1 Benefit Costs
- MSD Yearly Tier 2 Supplementary Benefit Costs
- MSD Tier 3 Benefits Amount received
- Older Adults Count
- Older Children Count
- Working Age Adults Count
- Young Children Count.

All costs and counts-related variables are computed on a 12-month basis over the 48 months leading to the application date (i.e. the 12-month period leading to the application/the 12 to 24 months period before application), leading to four variables computed for each category above. Two of the variables above were only computed over the 24 months leading to the application date, as indicated, for reasons of data availability.

The initial version of the work used variables describing counts and duration of events. Following feedback on the preliminary results, it was decided to use cost variables instead. After balancing, this ensures the two groups show similar prior costs, guaranteeing any observed difference in forward costs (i.e. after having received social housing support or not), do not reflect a prior condition.

3.1.1.4 Discarded variables

Other variables were considered useful to include in the model but issues were discovered with data quality, interpretability or high correlations with other variables. The most important of these are:

- Current Meshblock
- Current Territorial Authority Code
- Household Type
- No Location Preference Flag
- Preferred Location.

3.1.1.5 Outcome variables

The outcome/target variable used in the model was the application outcome for social housing, as provided by HNZ. This is a binary variable, and can be 'HOUSED' or 'OTHER EXIT'.

Applicants receive an 'OTHER EXIT' status when the circumstances change, affecting their eligibility or need for a social house, justifying their removal from the register. These applicants have not received a social house from HNZ, although it is possible they obtained housing from council providers, family or friends, went to the private market or were homeless. There was insufficient time to visit the pathways of the 'OTHER EXIT'. A separate piece of analysis would be needed to find out more about these different pathways. In the remainder of this document, these households are referred to as not 'HOUSED' by opposition to our treatment group.

These values can be recoded in a binary format, with '1' for 'HOUSED' and '0' otherwise. The dataset derived is roughly balanced on the number of applications 'HOUSED' and not 'HOUSED', so there was no need to perform any over/under-sampling processes while preparing the dataset for analysis.

3.1.2 Variable transformation and imputation

A few variables were transformed to obtain new representations thought to convey better information for estimating the probability of selection for social housing. This was often the case for important variables with skewed distributions. For example, the 12-month wages and benefits for the primary applicant were transformed into a logarithmic wage-benefit ratio.

In other cases, natural log transformations were used to handle skewed distribution of values. For categorical variables, collapsing of levels was performed in cases where the counts for some levels were considered too low.

In cases where data quality was poor or missing, alternative IDI data sources were searched to estimate the missing values. This was done while attempting to keep variable imputations to a minimum for the following reasons:

- In most variables with missing values, the number of missing values was too large to reliably perform imputations
- 'Missing-ness' could not be safely attributed to random processes, e.g. 'Preferred Location' and 'Location Preference Flag' variables had missing values but the missing-ness may not be completely random. This could be indicative of the pressing need for housing or homelessness that made the applicants compromise on location of the house, at least in a small subset of cases.

Therefore, imputations have only been applied to those variables deemed indispensable in terms of importance to the analysis or interpretation of the model. In other cases, the variable has been excluded from the model.

The newly transformed variables are:

- 12-month Household Income
- Age Category
- Application Main Reason
- Current Region Code
- Prioritised Ethnicity
- Wage to Benefit Ratio.

The business rules for the cases where transformations/imputations were performed on the dataset are detailed in Appendix D: Variable transformation rules.

3.1.3 Variable statistics

3.1.3.1 Collinearity

Multicollinearity refers to highly correlated variables. Such variables make working with logistic regression models difficult. On the other hand, some models, e.g. gradient boosting tree-based, are robust to the effects of multicollinearity. For this reason these models were chosen.

Testing was conducted using condition indexes and looking at correlations. Anything over 100 was considered to indicate severe multicollinearity.

The largest condition index indicated the intercept is highly correlated with itself. No reference coding was employed to the logistic regression model, so it is likely several of the reference levels refer to very small unknown categories. Instead, the largest condition index not affected by the design of the model was looked at. The value of this condition index is around 55, which indicates moderate collinearity only.

The highly correlated variables were the primary applicant's income and the household income, computed the year before application:

- 2 years before
- 3 years before
- 4 years before.

If only one person in the household is working, then the primary applicant's income is the same as the household income. Wages and salary (W&S) being relatively similar over a four-year window could also explain the correlation. It was decided to keep both of these

variables in the final model because they are important from a segmentation point-of-view, and it is important to ensure they are balanced between the two groups after application of the propensity scoring method. The effect of including correlated variables can be mitigated to an extent by using decision tree-based classification models for deriving the propensity score.

3.1.3.2 Summary statistics

Given the large number of variables, the summary statistics table can be found in Appendix C: Cohort descriptive statistics.

Pearson's chi-squared test was used to verify the relationship of the categorical variables with the outcome variable. Observations confirmed all categorical variables, except 'Gender', were correlated with the outcome of the application.

Table 2: Pearson's chi-squared coefficient for categorical variables

Features	Sample Size	Chi-Squared	Degrees of Freedom	p-value
Application Main Reason	21,822	500.79	18	<0.01
Current Region Code	21,822	375.8	16	<0.01
Prioritised Ethnicity	21,822	148.1	5	<0.01
Gender	21,822	0.93	1	0.33
Sole Earner Indicator	21,822	29.0	1	<0.01
Total Score	21,822	2065.6	3	<0.01

3.1.4 Understanding bias

When conducting propensity scoring, additional care must be taken with variables. Any propensity scoring analysis is vulnerable to biases resulting from the following:

- Differences in the likelihood of receiving social housing not accounted for by the model including:
    - Data not recorded or only recorded in free text fields (see discussion below)
    - Data is recorded but is not included in the Integrated Data Infrastructure (IDI) (see discussion below)
    - Failure of the propensity model to extract the relationships from the data.
- Repetition of the modelling and propensity matching using different assumptions until a desired outcome is achieved. To avoid this type of bias the modelling and matching approach were decided before the ROI was calculated. The modelling stage was also completed before the matching was conducted.
- Unbalanced treatment and comparison groups resulting from a loose matching process (see Section 3.3 Use of the model for propensity score analysis).

The primary source of bias unable to be addressed by propensity matching is when key information helping to determine the decision to allocate social housing to a client is not available to the model. For example, a tenancy manager may use information from a

discussion with an applicant to decide whether to allocate a house, without recording this fact, or only recording it in free text form.

To gauge the likely impact of bias due to unavailable data, MSD's Social Housing team and other MSD experts were consulted about what were the key drivers of receiving social housing support. They were also asked what data shortcomings they knew about that may influence the results. When explicitly asked about the 2005/06 cohort, the experts stated they were not sure whether these processes and drivers would have existed when HNZ was managing the Social Housing process. Consequently, these drivers should be treated with caution for our 2005/06 population.

Many of the primary drivers are available in the IDI (e.g. the total score), others are difficult to derive using IDI data (e.g. rheumatic fever), and some are not recorded (e.g. gang affiliations).

The key drivers for receiving social housing support identified by frontline staff and Subject Matter Experts (SMEs) were:

- The total score from the application process – this is not just the A, B, C, D priority[5] scale used in the model but also the sum of the scores for the individual sections. These scores are available to the model so there should be no source of bias from this driver.

- If one of the applicants has rheumatic fever – this is the only driver that will take an applicant to the top of the priority list. It is not included in the model so is likely to be a source of bias. The probable impact is the treatment group will have higher health care costs than the comparison group. This was not applicable to this test case population but, when looking at assessing the current social housing situation, it is necessary to be aware of this particular driver.

- The process of matching applicants to houses – this is done by social housing providers (rather than social housing staff), based on a list of up to 20 high priority applicants who match the provider's area. This process is not well understood and may vary significantly between providers. It is also not as well represented in the data available to the model. The provider may use criteria such as gang affiliations and smoking to select applications from the eligible list.

  It is unclear whether this will introduce bias, but providers are more likely to select 'easier' clients (less gang affiliations, non-smoking etc). The costs for these clients across the health care and criminal justice domains are likely to be lower than those not selected.

- Available social housing stock – applicants need to be matched to suitable social housing stock based on the number of adults and children listed in the application. There are also rules regarding the age and gender of the children, determining if they can share rooms. There can be large mismatches between the available stock and the applicants in a given region. For example, if there are more sole applicants than houses suitable for one person.

  The IDI contains limited data on HNZ stock and has poor coverage, so the propensity model cannot directly account for this driver. However, it does contain information with better coverage on the regional number of applicants and the number of children etc, so the model can indirectly account for some of the drivers regarding available

[5] Note: For future work people have not been assessed as Cs and Ds since a change to the needs assessment in 2011.

social housing stock. Matching on region should mitigate these issues. In addition, the second iteration of the test case involved constructing a year-quarter variable in an attempt to pick up fluctuations on the supply side.

3.2 Model training

The test case model has been built to estimate the probability of a household, which lodged an application for a social house, being granted one. The features considered as inputs (covariates) into these models are the ones described in Section 3.1, after pre-processing and transformation.

3.2.1 Models evaluated

Three methods were trialled to build the test case propensity model and performances of the resulting models were compared:

- Logistic regression
- Random forest
- Gradient boosting.

Logistic regression models are commonly used to build propensity scores. Their wide use is due to the ease with which models are trained and their high interpretability. They also allow for taking variable interactions into account when the interactions of interest are known in advance, usually due to subject matter expertise. However, as the dimensionality of the input of feature space grows, it becomes too computationally intensive to model and try all two or three way interactions between variables. The performances and accuracy of resulting models are consequently limited in the cases when single effects and two-way interactions are not enough.

On the other hand, Classification and Regression Tree-based (CART) methods provide a more efficient way to search and discover variable interaction of interest. They are also known to be quite robust to multicollinearity issues. These advantages come at the cost of a more complex tuning of the training process.

Random forest and gradient boosting models are two types of CART methods. The random forest algorithm involves independently building a (predefined) number of classification trees on a subset of the training data and a subset of the input space. Gradient boosting models differ from random forest in that trees are built sequentially, each tree being trained to fit the residual of the previous one, thereby progressively reducing the classification errors. While single decision trees are easy to interpret, random forest and gradient boosting models composed of multiple trees (sometimes hundreds), are far less so.

To ensure good generalisation of prediction and to prevent over-fitting, models were classically trained on a subset of the complete population and their predictive performances assessed on the remaining part. A 70/30 split was chosen.

In the initial population the two target classes (housed or not housed) are approximately balanced so no additional over or under-sampling was needed.

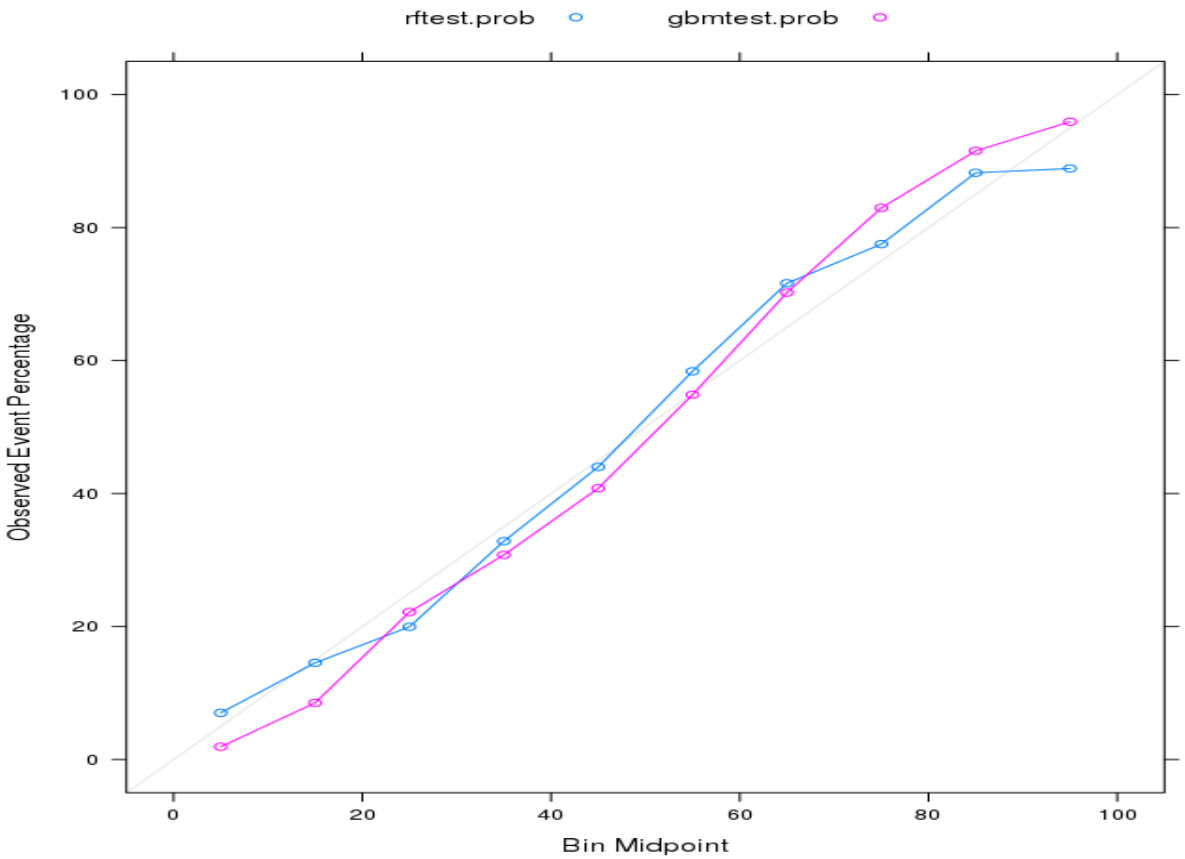
3.2.2 Comparison of model and selection

Both gradient boosting and random forest models showed close predictive performances and both outperformed the logistic regression one. Alongside these predictive performances, it is important predicted class probabilities effectively reflect the true underlying probability, that is, the actual occurrence observed in the training dataset.

This can be assessed using a calibration plot that shows the observed probability against the predicted class probability. Good models will exhibit calibration plots close to the Y=X line.

Figure 2 below shows calibration plots for both the random forest model and the gradient-boosting one. The plot was constructed using 10 points corresponding to 10 bins for the estimated probability, i.e. the left-most points (at x=5%) show the observed rates of occurrence of the event (applicant housed) among the subset of samples scoring between 0 and 10%. The right-most points show this rate among samples scoring between 90 and 100%. This plot shows the line for the gradient-boosting model is closer to the 45 degree line than the one for the random forest one. Estimated class probabilities reflect more closely the actual likelihood of the event under this model.

Figure 2: Two models fitted to the validation set: 'gradient-boosting trees' and 'random forest'



In this case, the gradient boosting method gave the best results and was consequently used to build the propensity score model.

The propensity score model was built for the propensity matching needed to balance the treatment and comparison populations in this test case. This model shows encouraging performances regarding the prediction of whether a given application for social housing support will be successful or not. Although it can be used to inform MSD and HNZ about the housing allocation process, it was not developed to replace the SAS. However, it does show the potential of such a model.

The training method used here was coded in a reusable manner, encouraging interested parties to reuse and extend it. The current model relies on a wide range of data from several

government agencies. It cannot be directly used on custom data sources outside the IDI environment without substantial redesign in terms of the variables supplied to the model. Such cross-sector agency data is only available through the IDI – consequently the model as it is could not be used outside the IDI environment, e.g. by MSD (frontline) staff.

3.2.3 Retained propensity score model

3.2.3.1 Training: Parameter selection and tuning of the model

The training of a gradient boosting model requires the tuning of a few parameters such as:

- maximum tree depth
- Learning rate (or step-size shrinkage)
- Row subsampling rate (setting the proportion of observations used for training)
- Column subsampling rate by tree (setting the number of features used for training)
- Number of trees built.

To identify and select the configuration leading to the best performing model, a set of values for each parameter has been set and a grid search performed where models were built for each combination of these values. Each model was trained following a 10-fold cross validation strategy combined with early stopping to prevent over-fitting. With each iteration of the training algorithm, the prediction error is monitored on the validation set and the training is stopped when this error shows a consistent increase over 10 iterations.

This strategy replaces the need to choose the number of trees built.

Table 3 shows the values tested for each parameter. The values leading to the best model obtained are in bold. Initial values reflect classic values for the parameters. The values for the maximum tree depth were selected based on the V-C dimension of the tree and dimensionality of the input space.

This value determines the maximum allowed n-way interaction between variables.

Table 3: Tuning of parameters for the gradient-boosting model

Parameter	Tested values
Maximum tree depth	2,3,4,**5**
Learning rate	0.01%, 1%, **10%**
Row subsampling	50%, 75%, **90%**
Column subsampling	**60%**, 80%
Number of trees*	**108**

* The number of trees is automatically set through early stopping.

Further details are in Appendix F: Tuning for the gradient-boosting model.

Performances were measured in terms of Area Under Curve (AUC) and misclassification rate. The models indicated in bold yield an AUC of 0.7476 and a misclassification rate of 31.46%.

Several configurations led to close performances. It is interesting to note the best four models obtained all had a depth of 5 and a step-size shrinkage of 0.1. One model gave a

better misclassification rate but a lower AUC. The configuration highlighted was selected as it performed consistently well over multiple training with different random seeds.

3.2.3.2 Relative feature importance

An interesting feature of gradient-boosting models is they allow estimation of the importance of input variables, measured as the relative contribution of each variable in the model predictions, compared to all other variables. These contributions are summed up across all trees making up the model to create the 'Gain' of the variable.

The total gain of all variables in the model sums to 1, as it is a relative measure.

The variable importance can also be calculated using other methods, like the 'cover' of a variable, defined as the number of records classified into leaf nodes based on the value of the given variable. This number is expressed as a percentage of all variables' cover metrics and also sums up to 1.

The feature importance was created for both random forests and gradient-boosting trees and found there was substantial overlap between the two models in terms of the most important variables. The most important variables can be found in Appendix E: Gradient-boosting model – variable importance.

Predictably, the assessment scores assigned by HNZ came out as the most important features for classification. Other variables of importance include:

- Wages and benefits received by the primary applicant and the household in the 12 months prior to application
- Region of the applicant, bedrooms required and the household size
- MoH-related claims
- Among the MSD benefits, duration spent on sole parent and sickness benefits are important predictors, along with tier 2 and 3 benefits
- Reason for application and ethnicities (only moderately important).

No further variable selection was performed on the basis of feature importance for the propensity matching, as academic literature advised against such strategy [see Brookhart *et al* (2006)]. Some researchers follow a statistical significance approach and start with a small subset before adding more variables until treatment groups are balanced. More information follows.

3.2.3.3 Variables and their incorporation into the model

Whether or not to use variable selection for propensity scoring is not clear. Some authors (e.g. Rubin and Thomas (1996)) advise against retaining only significant variables in the propensity score estimation, unless the variable is unrelated to the outcome of interest. Others employ a statistical significance approach and start small before adding more variables, until the treatment group balances. All 76 variables created in the gradient-boosting model have been used, meaning the estimated predicted probabilities used in the propensity matching are a scalar representation.

An earlier version of the analysis did not include cost variables in the model because the methodology for inflation and discounting the costs had not been finalised at the time the model was built. However, the risk of this is that differences between costs of those who are housed, which may have already existed before they received social housing, are seen. This has been corrected in the current version.

Behaviour can change when looking at the lead-up to an intervention. For example, in employment programmes there tends to be a fall in the proportion of clients off the main benefit in lead-up to participating in the intervention. This phenomenon is referred to as 'Ashenfelter's dip' and can be problematic for this type of analysis because it is difficult to find a suitable comparison group. To ensure similar people are found, costs were tracked costs at a detailed level (e.g. lab tests, pharmaceuticals, hospitalisations and so on), in -1, -2, -3, -4 year periods before the intervention (where data is available), to find someone who has a similar profile before they move into a social house.

3.2.3.4 Model performances

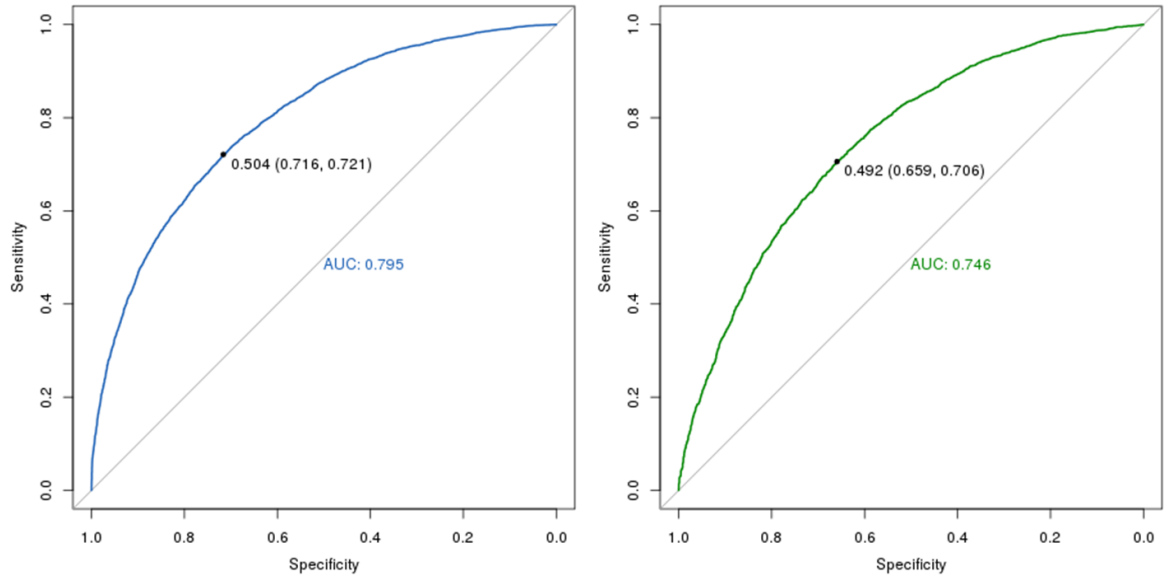
Once the models have been built, predictive performances of the selected model can be measured on the validation set (30% randomly selected observations set aside before training). The model exhibits a classification accuracy of 68%, which is quite close to the training accuracy of 72%. The AUC for the model on the validation set is 74.6%, compared to 79.5% on the training set (see Table 4 and Figure 3).

Table 4: Confusion matrix for gradient-boosting model

Confusion matrix (Training dataset)		Actual		Kappa	0.43
		Soc. housed	*Not Soc. housed*		
Model	*Soc. housed*	5377	2260	*Positive Predictive Value*	0.70
	Not Soc. housed	1997	5433	*Negative Predictive Value*	0.73
		Sensitivity	*Specificity*	**Accuracy**	0.72
		0.73	0.71		

Confusion matrix (Validation dataset)		Actual		Kappa	0.36
		Soc. housed	*Not Soc. Housed*		
Model	*Soc. housed*	2247	1146	*Positive Predictive Value*	0.66
	Not Soc. housed	1012	2356	*Negative Predictive Value*	0.70
		Sensitivity	*Specificity*	**Accuracy**	0.68
		0.69	0.67		

Figure 3: AUC measured for training (left) and validation (right) sets



3.2.3.5 Distribution of model variables by decile

The model was further evaluated on the basis of how well it was able to distinguish between housed and non-housed applications. The relationship between propensity to be housed and independent variable values are interpreted by looking at the behaviour of the variables across the deciles of risk scores predicted by the model.

Looking at behaviour of individual variables may mask the underlying interaction effects these may have on the model's propensity scores.

For brevity, only a selection of the plots are presented here. The remainder of the plots can be found in Appendix G: Selected covariates by risk decile.

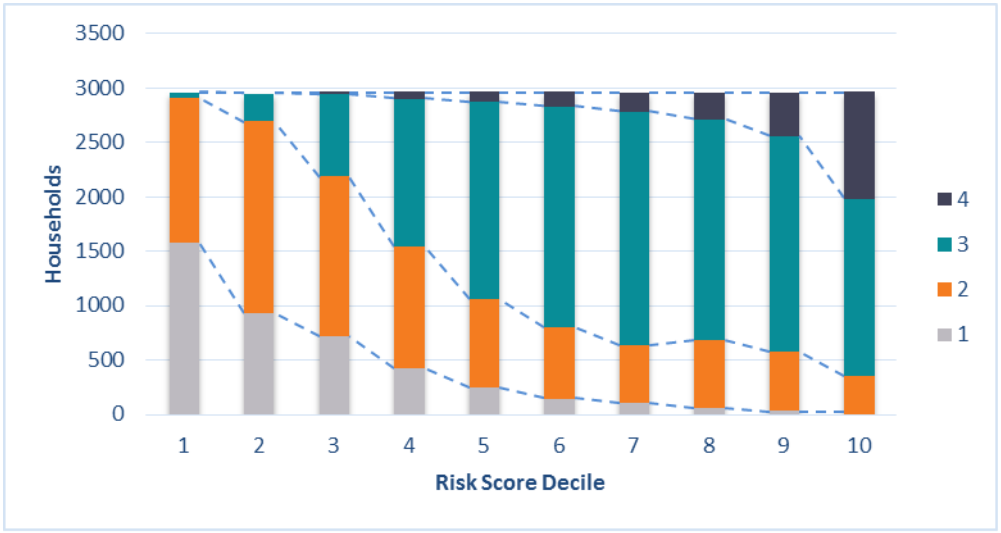
Sustainability, suitability and accessibility scores

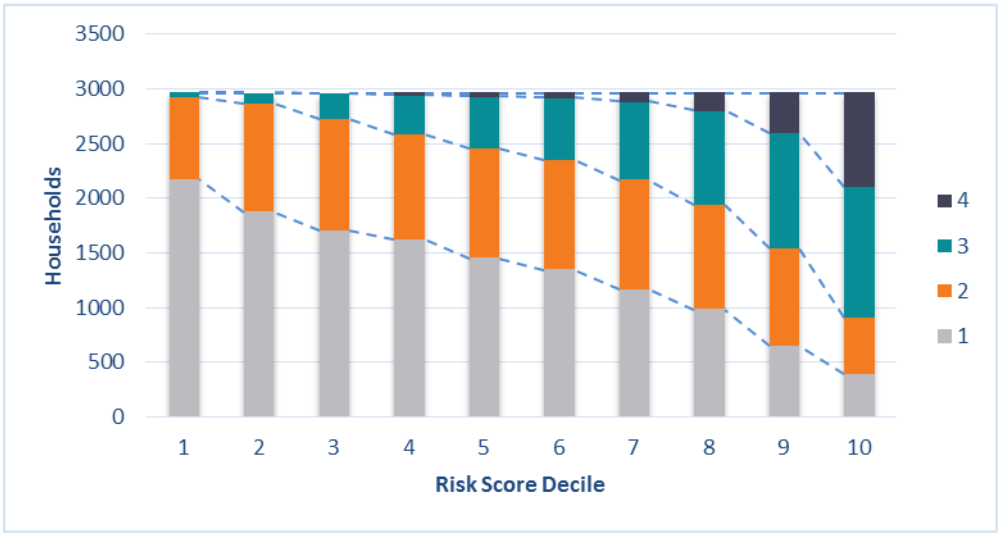
These three variables have been consistently marked as the most important across several model training iterations (both in gradient-boosting trees and random forests). The change in the score values across the predicted probability deciles follow. The graph given in Figure 4 clearly shows higher score values have higher probabilities for receiving social housing support, as expected.

Household size

An increase in household size is associated with a greater likelihood of placement in social housing. This can be observed from the model output, which shows an increase in the probability of being housed with larger household sizes (see Figure 5). Household size is also an important variable in the model for determining the probability of being housed. Consultation with MSD's social housing advisors revealed it is difficult to place sole applicants due to a shortage of suitable houses.

Figure 4: Sustainability (top), suitability (middle) and accessibility (bottom) scores by propensity score decile





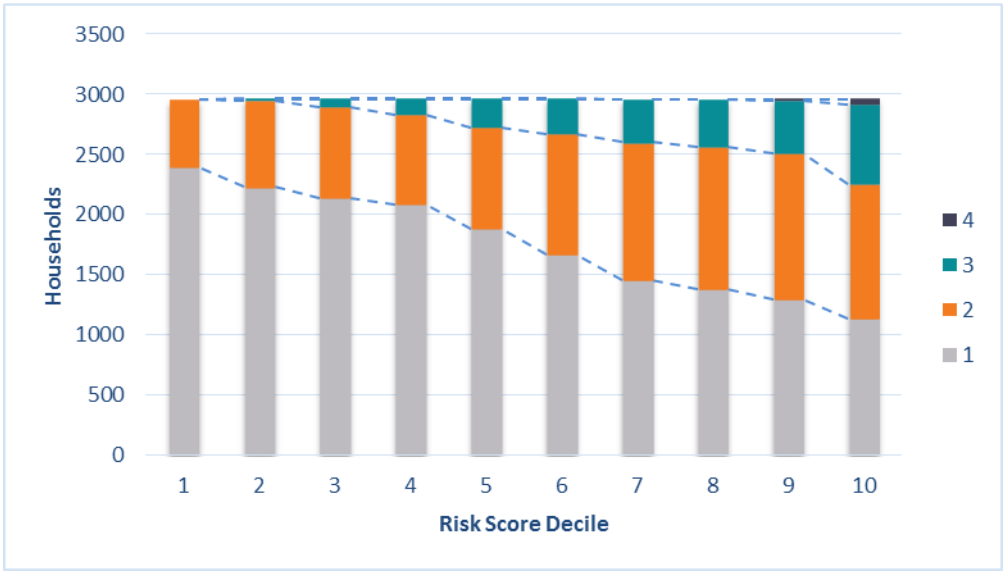
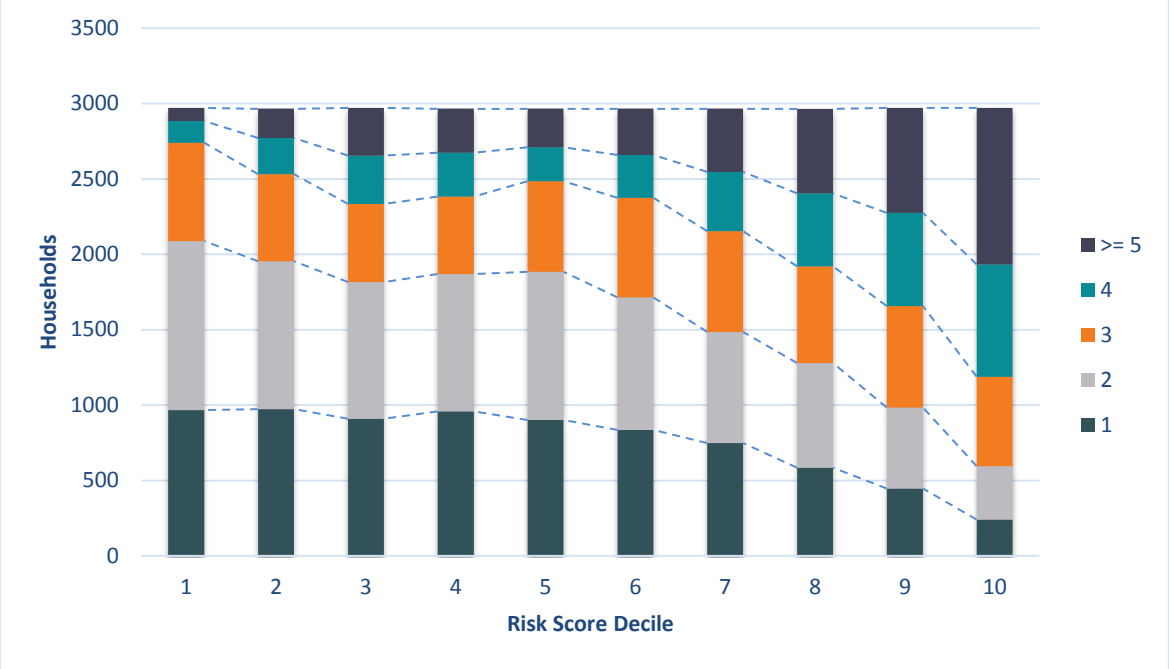


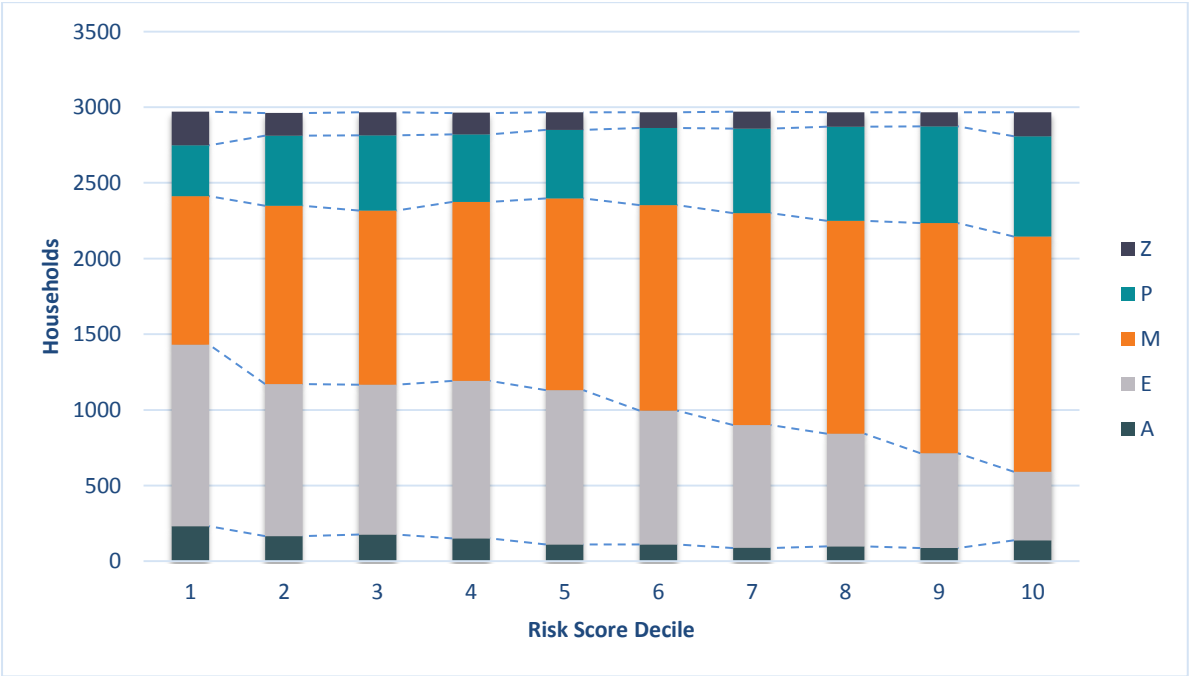
Figure 5: Household size by propensity score decile



Primary applicant ethnicity

The model shows the ethnicity of the primary applicant tends to have only a slightly observable effect on the predicted probability of receiving social housing support (Figure 6). This could be due to interaction effects with other variables, such as income levels or benefit receipt.

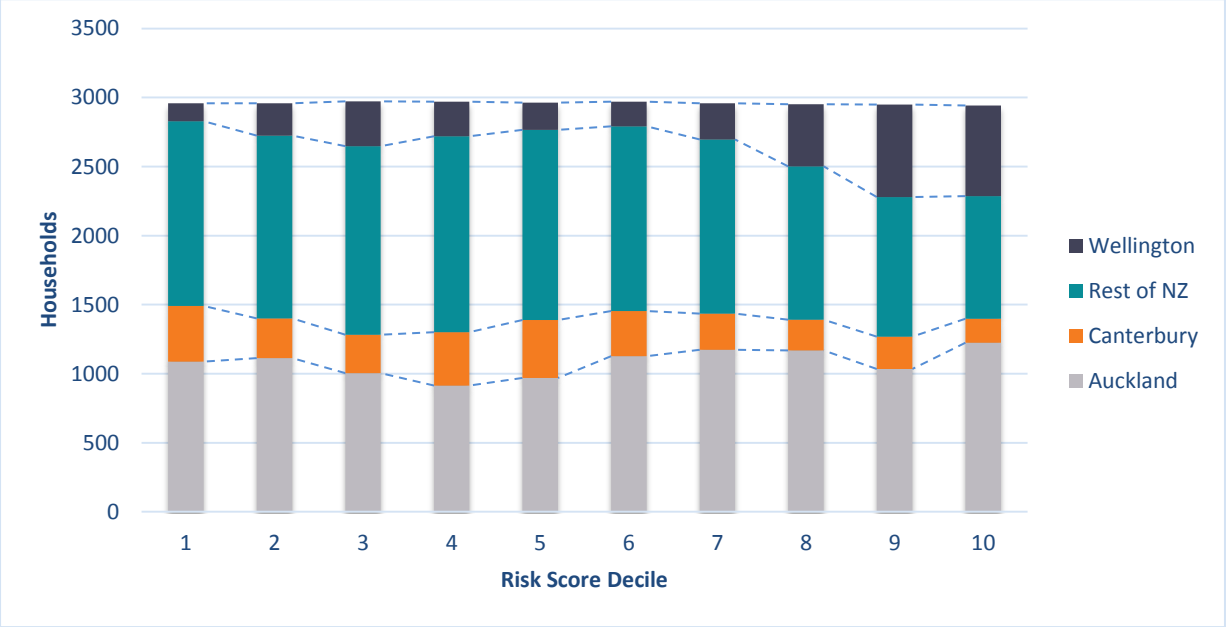
Figure 6: Primary applicant ethnicity by propensity score decile



Region of applicant

The model shows a higher probability of being housed if the applicant's region is Wellington (Figure 7). This could be explained by supply-side effects, like a greater supply of houses in this region. The effect in other regions seem to be more moderate.

Figure 7: Region of applicant by propensity score decile



3.3 Use of the model for propensity score analysis

3.3.1 Common support and balance

This section details the results of applying the propensity score-matching method introduced above. Following the conditioning on the propensity scores, it is expected there is no longer dependence on the covariates, enabling direct comparison of the groups and causal effects due to social housing.

Common support checks the overlap in propensity scores. Some might consider the use of propensity scoring for social housing as fundamentally flawed. It could reasonably be expected that those with greater need are more likely to receive social housing and, consequently, there would be no appropriate comparison group. It was found this is not the case. Some people whose need is not as great receive social housing because of excess supply of housing stock in their region. Sometimes those who have a great need are not housed because their regions have limited housing stock.

Observing a lack of common support, it is a useful finding in itself even though the two groups cannot be compared. Knowing those who participate differ substantially from the eligible population could be useful for designing interventions.

Figure 8 shows the distributions of scores for the two populations (treatment and counterfactual), before and after weighting, and that the weighted comparison population matches the treated one.

To make inferences about social housing, it is necessary to ensure the two groups are as similar as possible to each other. It is necessary to check the two groups are balanced by comparing standardised mean differences. Again, there is no gold standard for choosing an appropriate difference threshold – the literature suggests a threshold of 0.1 and up to 0.25 is a correct value.

The only variable exhibiting balancing issues is 'region' (see Figure 9). This is not surprising as this is one of the few supply-side variables available in the IDI. Having just a region variable is not sufficient to capture the variation.

As mentioned earlier, to help with the lack of supply-side variables, a quarter-year variable was constructed. However, without any further supply-side variables it would be difficult to get the standardised difference for region to be even smaller. The resulting implication is any break down looking at 'region' should be treated with caution, as the differences may not be due to social housing but rather differences in unobserved characteristics of the units in the two groups.

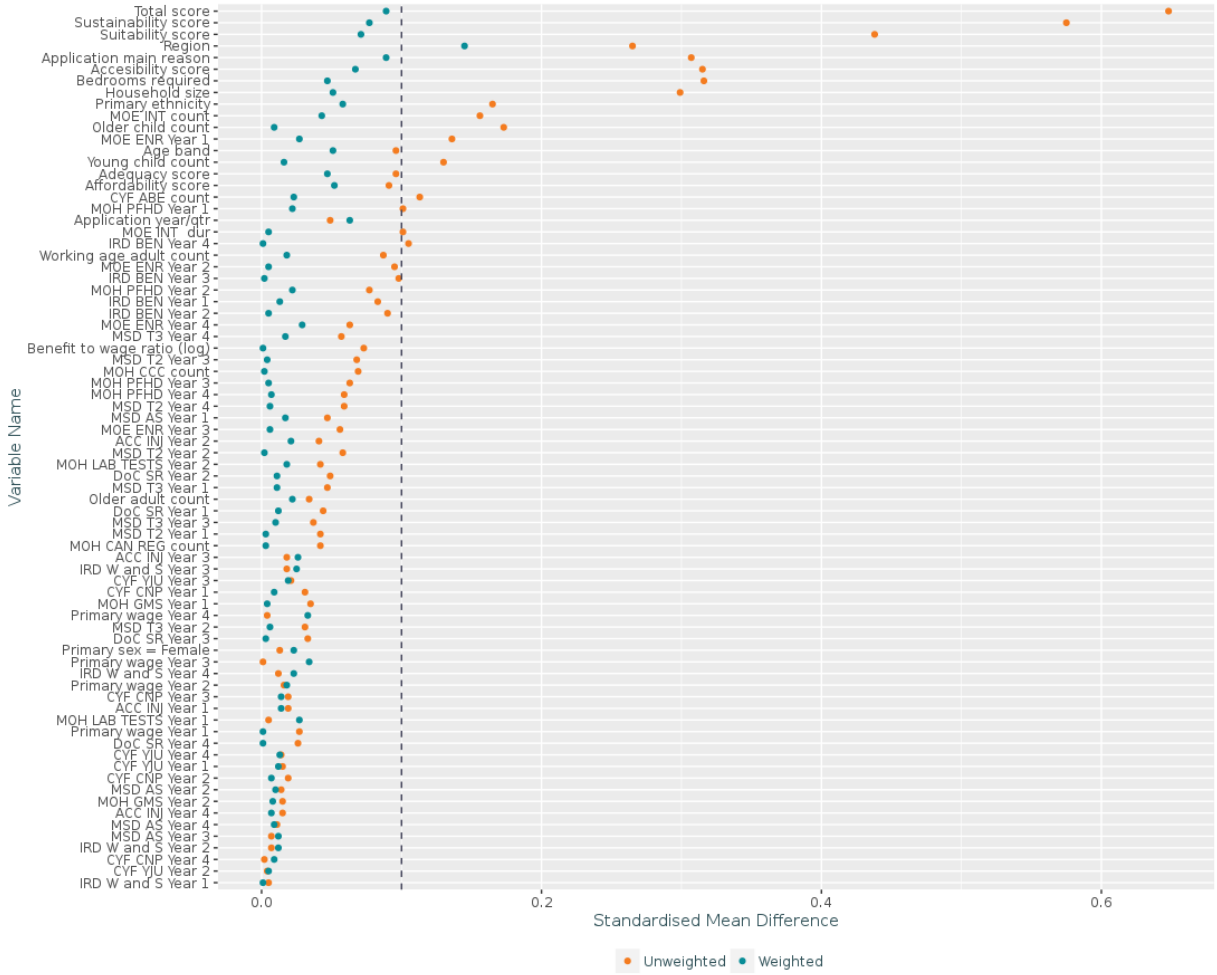
Figure 8: Comparative distribution of treatment and comparison populations before and after weighting

Often balance is not achieved the first time and iterations of the modelling process are necessary. If evidence still suggests the treatment groups to be significantly dissimilar, the model may have been wrongly specified or not all measurable covariates contributing to the bias have been accounted for.

Strategies helping to adjust confounding include:

- Adding more background variables or interaction terms to the model (for logistic regression models)
- Reducing areas lacking common support between the two groups
- Deleting extreme observations in either opposing group.

Figure 9: Balance plot showing standardised mean difference before and after weighting



3.4 Key lessons and future improvements

Propensity scoring is a key process in creating a counterfactual group. This ensures like applicants are being compared and differences seen in cross-sector spend is attributable to social housing. Possible future improvements include the following:

- It would be ideal if more supply-side variables were available to model the social housing process. This would ensure the model was better balanced and more

reliable inferences about cross-sector costs for people in different regions could be made.

- Feedback on the preliminary results highlighted it was not clear whether ATT were being looked at and why. Subsequently, this has been made explicit and weights, formulae and plots have been included. Future propensity scoring pieces of work will be treated similarly.

- One improvement could be to identify interactions from xgboost to build an easy-to-interpret logistic model. This would help with collinearity issues and could also improve interpretability of the model.

- The section on propensity scoring clearly highlights there is no consensus on how to evaluate propensity scoring, making it difficult to ensure a standardised process. The report illustrates different agencies use different approaches but, for social housing at least, it makes little difference to the results.

- An alternative way to tackle the idea of creating balance would be to use an optimisation rather than propensity scoring to minimise the imbalance between the covariates. This might be more foolproof but it could be computationally intensive.

- For common demographic variables, an exact match could be done, followed by a propensity score-based one. It is uncertain whether this would make a considerable difference.

- The bias section highlights a data limitation is the lack of supply-side information. The implication is that differences in costs due to social housing could result from uncontrolled factors. SIU made use of what was available, including region and a year-quarter variables, which should minimise this bias.

4 Calculating ROI

One of the stated goals of the analysis was to test the possibility of devising a measure of fiscal ROI for government spending on social policy. Section 2.5, Computing and monitoring costs, illustrated numerous costs can be monitored using the IDI. These government costs (and revenue in the case of tax paid), can be used to estimate a potential fiscal 'return' or benefit following social housing intervention, a decrease in monitored cost being taken as a benefit.

To complete the computation of the ROI, it is necessary to be able to compute the investment part, i.e. the amount of money paid by the government to provide the services.

Social interventions such as social housing have an impact on social and economic factors relevant to several government agencies.

When assessing the impact of such social policies, it is important to measure these broader impacts. One way is to measure through a fiscal lens, i.e. by monitoring the different fiscal costs and savings that occur for the relevant agencies – these costs and savings should then be considered alongside the cost of providing the services in the first place.

Although such a fiscal-only approach does not inform on all actual impacts on social outcomes (such as employment rate, education rate and attainment, child abuse rate), it does provide an indication of great importance to policy analysts and ministers who need information about what works, for whom and at what cost.

4.1 Principle: Constructing the investment

In 2005/06, there were two ways government spent money on social housing[6]:

- IRRS paid to HNZ who charged an IRR to their tenants
- Subsidising the capital providers have tied up in social houses.

When people invest in the private rental market, they receive returns from both rental cash flow and access to capital when the house is eventually sold. This is similar for social houses, however, the access to capital is limited as social housing places are not as liquid as private houses. It is more difficult to sell social houses on a short notice, consequently this capital is worth less money on the open market.

When the IRRS is based on a private market rent, capital gain is assumed to be at its full market value. This assumption results in an underestimation of the total cost of social housing. The real value of this capital is some, but not all, of its market value.

Underestimating the cost of social housing in calculating ROI is problematic for policy making as this would impact the comparison with alternative service providers such as CHP. This would also distort the perception of ROI across regions, since the relationship between rental yield and capital gain differs from region to region.

In assessing ROI, it was decided to use a method which strikes a pragmatic balance by using the weekly rents MSD is willing to pay, based on its experience of purchasing social housing on the open market. These figures reflect a rental yield and some compensation for the lack of access to capital experienced by social housing providers. When calculating the

[6] Note this has changed, as Government spend towards social housing services now includes flexible funding to Community Housing Providers, including up-front funding and operating supplements.

ROI, MSD's investment was determined by subtracting the IRR paid by each household from this rent. This avoided the need to compensate for the implicit capital subsidy built into HNZ properties charging market rent.

The specific details of how these are constructed are in Appendix H: More details on deriving the investment component.

To summarise, the HNZ tenancy snapshot table from the IDI is taken and a revised 'IRRS + capital' amount is attached to each house by: date (month/year), number of bedrooms and region (territorial authority (TA)), as supplied by policy.

All costs and revenues taken into account in this test case have been Cost Price Index (CPI)-adjusted and discounted so they reflect a net present value.

4.2 Difficulties of computing the investment

When benefits (returns) and costs (investments) amounts have been computed, the detailed calculation for the total ROI can be written simply as:

$$ROI = \frac{\left(\sum_{i=1}^{n_T}(revenue_i - cost_i) - \sum_{j=1}^{n_C} w_j.\left(revenue_j - cost_j\right)\right)}{investment}$$

Where:

w	The inverse predicted probability weight
n_T, n_C respectively	The number of applicants in our treatment and comparison groups
i	Households who applied for, and subsequently received, social housing support (treatment group)
j	Households who applied but did not receive social housing support (comparison group)
$cost$	All fiscal costs that can be attributed to the individual level across welfare (MSD), CYF (MSD), MoE, ACC, COR and MoH, in the IDI
$revenue$	Tax collected by government from wages and salaries
$investment$	The investment part, as detailed above.

However, in the test case some limitations inherent to our method of constructing our treatment and comparison group meant the calculation of this ROI was not straightforward. In particular, a closer look at the two groups revealed a leakage between the two – it appears some households in the comparison group eventually received social housing support during the six-year follow up period, thus incurring IRRS-related costs.

Similarly, some households in the treatment group left their social house to return to the private market during the follow up period and were subsequently paid AS and other housing-related benefits.

The question of how to take these cross costs into account must be given particular attention. Advice was sought from analysts from MSD's Investment Approach and Treasury's Analytics & Insights teams on the matter. The feedback received indicated:

- There is no universally-correct formula for ROI and different options are valid – albeit they represent different ROIs
- Group leakage should be addressed to derive an accurate and robust ROI figure. This requires careful thought and represents substantial work.

Discussions are underway with these parties to work towards a way of addressing these issues. In the meantime, it has been decided to focus on fiscal impacts (which inform the 'return' part), and not to report on a single ROI number. The details of fiscal impacts across government agencies are given in Chapter 5.

4.3 Limitations

The estimation of a fiscal-only ROI does not fully reflect the improvement (or degradation) of social and economic outcomes and wellbeing. This should as well be taken into account when evaluating social impacts.

This limitation is illustrated by the following example: the study found that households receiving social housing services were showing increased costs related to education (see next chapter). Although this had the effect of lowering the computed figure of the fiscal ROI, increased education costs (if due to longer school enrolment), can be linked to better social outcomes in the long-term (especially better employment rates). Similarly, higher health costs can be attributed to more frequent contact with health services which, if for prevention purposes, are actually more efficient on a pure cost basis. Such effects weren't visible within the timeframe examined here.

Several strategies could be followed to mitigate these limitations:

- Considering a long-term (even lifetime) forecast window on which both counterfactual and treatment groups are monitored. This would allow measurement of the effect of higher (short-term) education and health costs on (long-term) extra revenue (through a higher amount of tax paid), or lower future health costs.
- Monitoring actual outcomes for the individuals (such as effective school enrolment rates, sentencing rates, child abuse rates), to balance the negative image associated with higher costs.
- Splitting fiscal costs between investment/prevention and protection, with greater value placed on reducing costs in the latter category.

A single figure representing a fiscal ROI alone does not provide a complete picture of the impact of social interventions such as social housing. Reporting on a single number could be misinterpreted if the limitations detailed above are not acknowledged. Although improvement can always be made, SIU believes in the value of computing an ROI figure to inform policy analysts and will work towards deriving a meaningful and accurate number.

4.4 Key lessons and future improvements

The idea of calculating an ROI sounds like a simple concept but there were many complications to be considered:

- Where costs were not available they were derived based on the timeframes available. All these derivations could be improved. For example, when court case costs were sent to Ministry of Justice (MoJ) for business Quality Assurance, an alternative method was recommended that would be more appropriate to estimate the court case costs. Consequently, the MoJ court case costs were not included, as the new version of the costs was not available.

- Another shortcoming of some of the costs derivations was that an average effective tax rate of 13.39% was applied to the households' declared income to calculate government revenue from income taxes (due to time constraints).

- There is no cross-government agreed way to derive cost of capital. In the absence of an agreed (and considered) approach, it's inclusion in ROI becomes very arbitrary and difficult to determine whether social housing delivers value for money or not.

- Other aspects of the cost of social housing were not fully captured by IRRS – some may consider cost of capital as part of the investment. There was rigorous debate about whether or not to include the cost of a substitute (the AS in the denominator). Care needs to be taken when constructing the investment value as large changes in the denominator will result in large changes in the ROI.

5 Impact analysis and interpretation

Before reporting on the results, it is essential to recall the exact scope of the test case study: ATT analysis was conducted over the cohort of people who applied to HNZ for social housing over the period 2005/06. Consequently, the results describe the fiscal impact (how much money was spent/saved), when social housing support was granted over the given period. These results are not an estimation of the fiscal impact of all social housing services - only services provided by HNZ during the years 2005-2006 were taken into account. The results are not intended to represent the value of having a house in general.

5.1 Cross-sector spends results

Figure 10 illustrates cross-sector spend for those in social housing – orange represents less spending, while green indicates more spending. Table 5 shows the dollar figures in more detail. These results are discussed further in Section 5.2

Figure 10: Total cost difference by agency over the 6 year follow up period

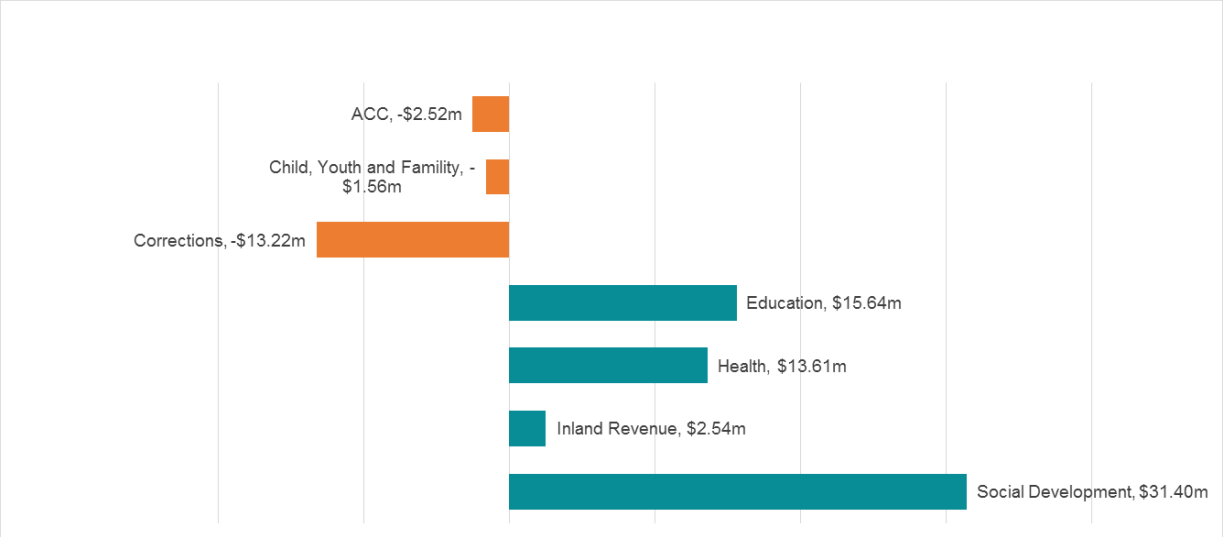


Table 5 reports the total (weighted) costs per agency. As stated, the numbers indicate trends at most. It is important to verify these results are statistically significant.

To assess whether these results are significant, confidence intervals need to be estimated to measure how likely the results are to vary due to sampling effects. An estimation based on a bootstrap strategy was performed – 1000 random samples (draw with replacement) of the treatment and comparison populations where generated and the total (weighted) fiscal costs were computed for each of these samples. In each run, new weights were computed using the existing propensity scores computed, and adjusted so the comparison group's sum of weights always equalled the total number of people in the treatment group.

Table 5: Total cost difference by agency over the six year follow-up period

Agency	Group	Weighted total cost	Absolute difference	Relative difference	Average difference per household
ACC	Comparison	$ 33,873,683			
	In social housing	$ 31,355,462	- 2,518,222	-8.0%	-$ 236.8
CYF	Comparison	$ 23,711,738			
	In social housing	$ 22,148,941	- 1,562,797	-7.1%	-$ 147.9
COR	Comparison	$ 65,370,904			
	In social housing	$ 52,150,022	- 13,220,882	-25.4%	-$ 1,244.2
MoE	Comparison	$ 252,249,856			
	In social housing	$ 267,886,004	15,636,147	5.8%	$ 1,467
IR[1]	Comparison	$ 21,359,162			
	In social housing	$ 23,896,925	2,537,763	10.6%	$ 239.3
MoH[2]	Comparison	$ 235,569,599			
	In social housing	$ 249,182,452	13,612,854	5.5%	$ 1,282.6
MSD[3]	Comparison	$ 847,331,083			
	In social housing	$ 878,735,618	31,404,535	3.6%	$ 2,919.4

[1] Includes Paid Parental Leave and Student Allowance
[2] Excludes PRIMHD and PHA costs
[3] Includes tier 1, tier 2 and tier 3 benefits payment only. All accommodation-related costs (AS and IRRS) are excluded.

Table 6 reports on the 95% confidence intervals estimated using the bootstrap strategy.

Average fiscal impacts (costs or benefits) per household are presented. The total costs can be obtained by multiplying these numbers by the size of the cohort (10,612). Confidence intervals including the value 0 indicate the corresponding result are not statistically significant.

In addition to government agency costs, the amount of W&S declared by households over the period was monitored in the IDI. These W&S allow the estimation of the taxes paid by households to IR[7] to be derived, constituting the revenue part in the return. The impacts on taxes are given in Table 6. The negative value for this measure indicates households receiving social housing support declared less income than households not receiving support.

5.2 Interpreting the results

The confidence intervals reported above show, at the agency level, the results are statistically significant only for COR (reduced costs following social housing), MoE (increased costs), and MSD (increased costs).

[7] A flat rate of 13% was applied on the declared W&S to estimate this tax.

Table 6: Average fiscal impacts per household with confidence intervals (by agency)

Agency	Average impact per household	Confidence interval – Low value	Confidence interval – High value	Statistically significant
ACC	-$ 236.8	-923.4	370.6	N
CYF	-$ 147.9	-929.9	547.9	N
COR	-$ 1,244.2	-2,121.8	-407.7	**Y**
MoE	$ 1,467.8	198.5	2,716.5	**Y**
IR	$ 239.3	-16.0	484.3	N
MoH	$ 1,282.6	-331.9	2,807.5	N
MSD	$ 2,919.4	1,024.8	4,830.2	**Y**
Tax (W&S)	-$ 556.6	-1,038.4	-78.5	**Y**

Figure 11, Table 5 and Table 6 are summarised in tabular format in Table 7. The shaded arrows represent non-significant results at the 95% confidence level.

Possible interpretations of the significant results only are discussed below. All costs and savings mentioned are computed over the six year follow-up period.

Table 7: Cost difference direction as a result of social housing

Agency	ACC	CYF	COR	IR	MoE	MoH	MSD	Revenue
Cost difference direction	↓	↓	↓	↑	↑	↑	↑	↓

Corrections

Overall, the cost for sentencing and remand is less for those who receive social housing. The difference in CORs' spend between social housing tenants housed (in 2005 or 2006) and the comparison group is a decrease of $13m, or a 25% saving.

A closer look at actual events in the follow-up period shows a similar number of households have some kind of interaction with Corrections' in the treatment and comparison groups. However, the detailed count of events reveals that, while the two groups show close numbers of community service and home detention sentences (and close associated total duration of events), the number of remand and prison sentences is substantially smaller for the housed group, compared to the comparison one.

Education

Overall, the cost for education is higher for those who receive social housing compared with the treatment group. The difference in MoE spend between social housing tenants housed (in 2005 or 2006), and the comparison group, is an increase of $15.6m, or a 6% increase in spend.

Further analysis established this is because, on average, children and teenagers in social housing stay in education longer, compared to like households not in social housing.

Although this is reported as an extra cost to government, it can be expected this extra time spent in school will eventually lead to better outcomes for the children in households receiving social housing support.

MSD (benefit receipt)

The difference in MSD benefit receipt between tenants in social housing (in 2005 or 2006), and the counterfactual group is an increase of $31.4m, or a 3.6% increase in spend. Benefit payment towards households receiving social housing support increased overall. However, it is interesting to note that while tier 1 benefit increased substantially (around $3,600 more per household over the six years follow-up period), tier 2 benefit payments (excluding AS) decreased by around $800 – see Table 8.

However, this does not represent the true impact of the social housing services provided over the period for MSD, as the resulting decrease in AS and increase in IRRS are not reflected in this number.

W&S

A reduction of $5.9m was observed in the total amount of tax on W&S paid between the treated group and the counterfactual group over the six year follow-up period, indicating a lower amount of declared W&S for the treatment group. This could suggest social housing support presents a disincentive towards working, as it may result in a loss of advantages. However, establishing there is a causal relationship would require further research.

5.3 Detailed results per subject areas

As discussed, at the agency level only three expenditure items show statistically significant differences between the treatment and the comparison group. However, reporting on these results at a more detailed level revealed more differences between the two groups.

In particular, it revealed a decrease in GMS (General Medical Subsidy) as well as PRIMHD (mental health-related) costs, and an increase in the Student Allowance (IR/STU). In terms of benefit-related costs, it appeared that tier 1 and tier 3 spends are increased on households receiving social housing, while tier 2 spends (excluding AS-related) decreased.

5.4 Key lessons and future improvements

One of the key goals of this test case was to calculate a fiscal ROI on a social sector intervention, if possible. If so, the next goal was to develop a reusable methodology and to understand the limitations of such a methodology.

- Results show that a fiscal-only ROI is not enough when assessing the impact of social sector interventions, but that looking at the breakdown of the cross-sector spend is useful. Feedback suggested a more useful breakdown would be to look at spending on protection versus spending on investment. This would make it easier to interpret agencies with a mix of good and bad spending, such as MoH (preventive vs. reactive care). This is certainly a worthwhile future improvement but would take some time to refine.

- As anticipated, the results raise more questions than answers. It is difficult to derive insights for outcomes tied to individuals, such as a history with Corrections. For example, a large decrease in Corrections' costs at the total level was seen. It would be interesting to see if this was due to an overall decrease in time spent in Corrections' services across the board, or whether it was a decrease in a small number of people with very high Corrections-related costs. As the weights were derived at a household level, this was not possible. Household weight can also not be used to answer specific questions about cost drivers.

Table 8: Average fiscal impacts per household with confidence intervals – by subject area

Dept/agency	Subject area	Average cost difference ($)	Confidence interval Low value	Confidence interval High value	Statistically significant
ACC	CLM	- 267.80	- 909.02	373.43	N
	INJ	38.75	- 34.63	112.14	N
CYF	CNP	- 170.07	- 887.16	547.02	N
	YJU	23.67	- 10.54	57.88	N
COR	S&R	- 1,250.37	- 2,083.88	- 416.85	**Y**
IR	PPL	- 36.95	- 82.19	8.29	N
	STU	272.58	35.84	509.31	**Y**
MoE	ENR	1,484.77	222.68	2,746.85	**Y**
MoH	B4S	- 0.16	- 0.90	0.59	N
	GMS	- 17.83	- 25.50	- 10.17	**Y**
	NNP	630.73	- 567.79	1,829.25	N
	PFH	1,373.86	- 169.71	2,917.44	N
	PHA	253.55	- 94.18	601.29	N
	PRI	- 1,053.80	- 1,808.55	- 299.04	**Y**
	TES	- 18.39	- 29.60	- 7.17	**Y**
MSD	T1	3,607.07	1,971.58	5,242.55	**Y**
	T2	- 863.43	- 1,313.83	- 413.02	**Y**
	T3	203.38	129.52	277.24	**Y**

6 Future work

The work presented was intended to be exploratory. Among other goals, it aimed to:

- Demonstrate the value of taking a social investment approach to evaluating policies
- Establish the foundation for a robust data-driven analysis method and identify any limitations in doing this
- Generate interest for future collaboration between involved parties.

Several limitations have been identified and outlined in this report, particularly the difficulty in devising a meaningful ROI number in the context of social housing.

Some limitations of the current method used to build and monitor the cohort of interest impact the validity of the results reported:

- The unit of analysis selected for monitoring was at the household level, not the individual. These households were followed as they were defined in their application for social housing. It is known and accepted that household composition (i.e. the individuals who reside in a social house), tends to change over time. This change in composition has not been accounted for in this analysis. This is a substantial limitation.
- Households granted social housing were monitored over the six year follow-up period without taking into account the effective tenure. While some households would have stayed in the house for the whole period, a substantial proportion would have left (e.g. to return to the private housing market), after a varying period. The analysis did not discriminate between those households staying in the house for a few months, a few years or the whole six years.
- Similarly, some households in the comparison group may have received a social house at a later date, within the six-year follow up period. This has not been studied in detail in this analysis.
- Results have been reported averaged over the whole (treated) group, without discriminating between profiles and characteristics of households. i.e. single tenants and extended families are accounted for similarly in the test case.

These last two points have the adverse effect of flattening the results by averaging and masking the discrepancy in what may potentially show a wide spectrum of impacts. It is natural to expect households with different characteristics and lengths of tenure would exhibit different outcomes, benefits or costs from receiving social housing assistance. These potential differences have been masked by the approach taken.

An important goal of social investment is to answer the questions:

- What works?
- For whom?
- At what cost?

Because of the limitations detailed above, the ‘for whom’ question has not been addressed sufficiently in this test case. To rectify this, one option would be to perform a data-driven segmentation exercise to identify a set of different profile types. The methods presented here could be then reproduced rapidly on each of the segments to measure the returns and

behaviours exhibited by each. This would be a first step towards discriminating results by profiles. This work is currently underway.

A more detailed but complex method would be to build a predictive model estimating the monitored cost (either total or per item), with respect to the detailed characteristics of the household. While a Generalised Linear Model (GLM) would have the effect of highlighting which characteristics impact these costs, it may not be powerful enough to lead to good prediction abilities. More complex models, such as Neural Networks, can be used to overcome this issue – with the adverse effect of only being used for predictions, not explanations.

If taking such an approach, it would make sense to take the length of tenure into account at the same time. If this is not done, the training set (showing monitored costs with regards to household characteristics only), will likely show too large a variation to allow an accurate model to be built.

The work needed to properly address the issue of tenure is substantial. The information would need to be extracted from the available data, which is not always of sufficient quality.

Importantly, this test case highlighted the limitations of measuring social outcomes through a fiscal lens only and over a short period. This is illustrated by the increased spend in education for social housing tenants. On a purely fiscal point, this has the effect of lowering the ROI. In reality, this may correspond to a better social outcome for children (better education resulting in a better employment rate), as well as a greater government revenue through taxes collected in the long-term. Similarly, the observed decrease in spend towards sentencing and remands does not say much about the actual impact of social housing with regard to offending rates.

To overcome this limitation, it is necessary to go further than a solely fiscal impact analysis and to study the actual outcomes in greater detail. This would paint a far more accurate picture of the effective impact of social housing – and of any future social policy questions. Ideally, the monitoring of such outcomes could be coupled to a proper 'human-centric' framework, producing an agreed measure of social wellbeing. It is worth noting the SIAL, created by SIU as part of this project, will facilitate the monitoring of outcomes.

SIU and collaborating agencies are already working on this. Treasury's Analytics and Insights team is working towards defining and monitoring outcomes. Projects have also been commissioned in several agencies (MSD/Ministry for Vulnerable Children, Oranga Tamariki (MVCOT), Treasury and SIU), on the definition of a well-being framework.

In social housing policy, a larger piece of work involving several agencies is starting under MSD's leadership, with the SIU's involvement. The idea of social investment has been embraced and will be part of the approach.

7 Recommendations

The Social Housing Test Case was developed in partnership with iMSD and MSD's Social Housing Policy team to illustrate a rigorous evidence-based approach to social investment.

A fiscal ROI was a useful starting point for evaluating a social sector intervention, as it was the only information available. Many additional benefits are likely to reside in the social category via improved quality of life. There were also non-fiscal outcomes of more interest, such as: *Does social housing help educational attainment?* These would require further research but this test case was an important first step.

Parts of the methodology can be standardised, while other parts cannot. SIU has created reusable standardised events tables (SIAL) for future analysis. These are being tested by a small number of teams external to SIU. The methodology of propensity scoring depends on how easy it is to construct a counterfactual group.

Planning for the second test case on Mental Health and Addictions has already indicated the difficultly in constructing a counterfactual group for mental health. However, there are other areas where it may be possible to use this methodology, such as education training and programmes aimed at getting people back into work.

Because social sector problems are inherently more difficult than simplified statistical problems, there are limitations to studying them. SIU has attempted to share their decision-making by incorporating a decision log process in Appendix I: Decision Log.

In addition, limitations have been clearly outlined at the end of each main section of this report.

A list of caveats is given in Appendix J: Caveats, limitations and assumptions.

8 References

Austin, P.C. (2011). An introduction to Propensity Score Methods for Reducing the Effects of Cofounding in Observational Studies, in Multivariate Behavioral Research, 46:399-424

Austin, P.C. & Stuart E.A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies, in Statistics in Medicine 34:3661-3679

Brookhart, M.A., Schneeweiss, S., Rothman, K.J., Glynn, R.J., Avorn, J., Stürmer, T. (2006) Variable selection for propensity score models, in American journal of epidemiology 163 (12):1149-1156.

Rubin, D. B., Thomas, N. (1996) Matching using estimated propensity scores: relating theory to practice, in Biometrics 52 (1):249-264

9 Abbreviations and glossary

Abbreviations

ACC	Accident Compensation Corporation
AMSTAR standard	Tool to assess the methodological quality of systematic reviews
AS	Accommodation Supplement
ATT	Average Treatment Effect on the Treated (group)
AUC	Area Under Curve
CART	Classification and Regression Tree-based
CBA	Cost Benefit Analysis
CHP	Community Housing Provider
COR	Department of Corrections
CYF	Child, Youth and Family
DHB	District Health Board
DIA	Department of Internal Affairs
ERP	Estimated Resident Population
ESOL	English for Speakers of Other Languages
F&C	Forecasting and Costing
GAM	Generalised Additive Model
GBT	Gradient Boosted Tree
GLM	Generalised Linear Model
GMS	General Medical Subsidy
GCV	Generalised Cross Validation
HNZ	Housing New Zealand
IDI	Integrated Data Infrastructure
iMSD	Insights MSD
IPTW	Inverse Probability of Treatment Weighting
IR	Inland Revenue
IRC costs	International Rescue Committee – a cost analysis methodology
IRR	Income-Related Rent
IRRS	Income-Related Rent Subsidy
MoE	Ministry of Education
MoH	Ministry of Health
MoJ	Ministry of Justice
MSD	Ministry of Social Development
MVCOT	Ministry for Vulnerable Children, Oranga Tamariki

NGO	Non-Governmental Organisation
NNPAC	National Non-admitted Patient Collection
PFD	Publicly funded hospital discharges
POL	New Zealand Police
PRIMHD	Pronounced ‘primed’ - MoH mental health activity and outcomes data
PSM	Propensity Score Matching
PTCE	Person Time Cost Event
RCT	Randomised Control Trial
ROI	Return on Investment
S&R	Sentencing & Remands
SAS	Social Allocation System
SIAL	Social Investment Analytical Layer
SIU	Social Investment Unit
SME	Subject Matter Expert
SMS	Maryland Scientific Methods Scale
TA	Territorial Authority
TAS	Temporary Additional Support
UID	(Statistics NZ) Unique Identifier
W&S	Wages and Salaries
WIES	Weighted Inlier Equivalent Separations
YU	Youth Justice

Glossary

Area Under Curve (AUC)	Used in classification analysis to determine which of the used models predicts the classes best
Average Treatment Effect on the Treated group (ATT)	Measuring the Average Treatment Effect on the Treated, that is, the difference in outcomes due to receiving treatment, corrected for all other confounding factors
Classification and Regression Tree (CART)	Umbrella term for Classification and Regression Tree-based methods provide a more efficient way to search and discover variable interaction of interest. Classification tree analysis is when the predicted outcome is the class to which the data belongs
Cohort	A group of subjects who share a defining characteristic, typically subjects who experienced a common event within a selected time period
Collinearity (and multi-collinearity)	A phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, meaning one can be linearly predicted from the others with a substantial degree of accuracy
Counterfactual (group)	A second group for observing and comparing results to those expected if an intervention had not taken place
Covariates	Any of two or more random variables that are possibly predictive of the outcome under study
Decile	Any one of nine numbers that divide a frequency distribution into 10 classes, each containing the same number of individuals
Estimated Resident Population	An estimate of all people who usually live in New Zealand at a given date
Fiscal	Relating to the public treasury or revenues, or financial matters in general
Generalised Additive Model (GAM)	A flexible generalisation of ordinary linear regression allowing for response variables that have error distribution models other than a normal distribution
Integrated Data Infrastructure (IDI)	Large (Statistics NZ) research database containing microdata about people and households. Data is from a range of government agencies, Stats NZ surveys and non-governmental organisations
Liability	Monies owed; debts or pecuniary obligations as opposed to assets; liabilities as detailed on a balance sheet, especially in relation to assets and capital
Logistic regression models	Used to build propensity scores; their wide use is due to the ease with which models are trained and their high interpretability

Propensity score matching	Statistical matching technique estimating the probability of an individual receiving a treatment based on a set of identified characteristics
Randomised Control Trial (RCT)	A study in which subjects are allocated by chance to receive one of several clinical interventions. One of these interventions is the standard of comparison or control
Return on Investment (ROI)	A ratio of net benefit to cost; the amount of return on an investment relative to the cost of the investment, expressed as a percentage
Segment	Group of people who share a common set of characteristics
Spine	Term given to IDI data once it has been linked at the individual level.
WIES	Weighted Inlier Equivalent Separation (WIES) is a cost weight adjustment used by MoH for time spent in hospital.

Appendix A Descriptions of variables used

This is a list of variables used in the analysis with fuller descriptions of what each variable represents. They are broken down into HNZ application variable descriptions, primary applicant characteristics variable descriptions, household characteristics variable descriptions and discarded variables.

Table 9: HNZ application variable descriptions

Features	Type	Description
Accessibility Score	Numeric	HNZ Scoring Criteria based on application evaluation. This variable describes the ability to access and afford suitable and adequate alternative housing as a result of discrimination, lack of financial means to move, and availability of adequate housing on the private market. '1' represents low need and '4' represents higher need.
Adequacy Score	Numeric	HNZ Scoring Criteria based on application evaluation. This variable describes physical conditions and availability of essential facilities of the applicant's existing house. '1' represents low need and '4' represents higher need.
Affordability Score	Numeric	HNZ Scoring Criteria based on application evaluation. This variable describes the ability to afford alternative housing in private market. '1' represents low need and '4' represents higher need.
Application Main Reason	Categorical	Reason stated by applicant for making the application for social housing.
Bedroom Count Required	Numeric	Number of bedrooms required for the social house as determined by HNZ based on applicant's needs.
Current Region Code	Categorical	Current region of the main applicant. If absent in the HNZ data, this was estimated from the last notified address before application date from Statistics NZ address notification data. Almost one-third of the current region values were missing from the HNZ data for the 2005/06 cohort, which had to be estimated from the last notified address values.
Size of Household	Numeric	Size of household as determined by HNZ in the application. This value can also be estimated from the number of individuals under an application but currently we use the figure provided in the application directly.
Suitability Score	Numeric	HNZ Scoring Criteria based on application evaluation. This variable describes overcrowding, lack of security of tenure of current house and medical/personal needs. '1' represents low need and '4' represents higher need.
Sustainability Score	Numeric	HNZ Scoring Criteria based on application evaluation. This variable describes financial management difficulties, difficulties in social functioning and lack of social skills. '1' represents low need and '4' represents higher need.

Total Score	Categorical	Total Score assigned to the application on the basis of the 5 individual scores. The score may also have subjective influences which may reflect the case manager judgements – so this variable is included in the superset of features. The attribute has 4 levels – A, B, C, D – where 'A' is the highest priority.

Table 10: Primary applicant characteristics variable descriptions

Features	Type	Description
12-month Benefits (primary applicant)	Numeric	Total MSD benefit receipt that the primary applicant received in the 12 months before the application date. This may include first tier benefits received by the primary applicant (including Working for Families Tax credits), ACC claims, Pensions, Paid Parental leave and Student Grants.
12-month Wages (primary applicant)	Numeric	Total wages earned by the primary applicant in the 12 months before application. This only includes the wages and salaries received by the person through employment.
Age	Numeric	Age (in years) of the primary applicant as at the application date.
Age Category		Same as above, but grouped into categories. 0-19, 20-35, 36-65 and >=66 were used.
Marital Status	Categorical	Marital Status of the primary applicant as at the application date. This is based on DIA records of marriages and civil unions.
Prioritised Ethnicity	Categorical	The prioritised ethnicity of the primary applicant based on Statistics NZ records.
Gender	Categorical	The gender of the primary applicant based on Statistics NZ records.
Sole Earner Indicator	Categorical	Indicates whether the primary applicant is the sole earner in the household, among all others included in the application. The indicator is based on 12 months of household income data prior to the application date. This is determined on the basis of whether the household income is the same as the primary applicant's income.
Wage to Benefit Ratio	Numeric	The Wage to (MSD) Benefit ratio of the primary applicant, for the 12 months preceding the application date. A natural log has been applied to this attribute to address the skewness of the distribution. $$ln\_prim\_wg\_ben\_ratio\_12mnth = \log_e(1 + primary\_wage\_12\_mnth) - \log_e(1 + primary\_benefit\_12\_mnth)$$ If either the wage or the tier 1 benefit amounts of the primary applicant are unavailable in the IR data then the wage/benefit is considered to be $0.00. All IR records with W&S are considered to be part of Wages, and 'BEN' (Benefits), 'CLM' (ACC compensation Payments), 'PEN' (Pensions), 'PPL' (Paid Parental Leaves), 'STU' (Student Loans) are considered as benefits.

In Table 11, all cost and payments variables are computed over 12 months periods, in the four years leading to the application date. That is to say, for each of the costs detailed in the table, four variables are created, labelled <*variable_name*>_Y*i*, for *i=1…4*, that represent the total costs over the 12 months leading to the application (<*variable_name*>_Y1), between the 24th to 12th months prior to the application (<*variable_name*>_Y2), between the 36th and the 24th months prior to the application (<*variable_name*>_Y3), and finally between the between the 48th to 36th months prior (<*variable_name*>_Y4).

Table 11: Household characteristics variable descriptions

Features	Type	Description
Household total wage and salaries	Numeric	Total household declared wage and salary for the applicant household by 12 months periods, over the four years prior to application, calculated as the sum of all wages received by all members included under the application.
Household total tier 1 benefits payments	Numeric	Total amount paid towards tier 1 benefits, summed up for all members of the household in the four years before application date, by 12 months period. Note: sourced from IR data.
Household total tier 2 benefits payments	Numeric	Total amount paid towards tier 2 benefits, excluding AS, summed up for all members of the household in the four years before application date, by 12 months periods.
Household total Accommodation support payments	Numeric	Total amount paid AS, summed up for all members of the household in the four years before application date, by 12 months periods.
Household total tier 3 benefits payments	Numeric	Total amount paid towards tier 3 benefits, summed up for all members of the household in the four years before application date, by 12 months periods.
CYF Abuse Count	Numeric	Count of substantiated abuse events that the children in the applicant household were victims of, in the four years before the application date. Based on CYF data.
CYF – CNP total costs	Numeric	Total costs related to Care and Protection (CNP) events, summed up for all members of the applicant household in the four years prior to application date, by 12 months periods.

Features	Type	Description
CYF – YJU total costs	Numeric	Total costs related to Youth Justice (YJU) events, summed up for all members of the applicant household in the four years prior to application date, by 12 months periods.
COR S&R Cost	Numeric	Total costs related to sentencing and remand events recorded against the individuals of the household as perpetrators, in the four years before application date, per 12 months periods.
Accidents/Injuries – ACC Claims Cost	Numeric	Total costs related to accidents/injuries claims summed up for all members of the applicant household in the four years prior to application date, by 12 months periods. Based on ACC data.
MoE Student Enrolment Cost	Numeric	Total costs related to school enrolments summed up for all members of the applicant household in the four years prior to application date, by 12 months periods.
MoE Student Interventions Count	Numeric	Count of student intervention incidents summed up for all individuals included in the application.
MoE Student Interventions Duration	Numeric	Total duration (in number of days) of Student Interventions for all individuals of the application in the four years prior to application date. The intervention includes Alternative Education, Suspensions, Stand-downs, Truancy Services (Non-Enrolled and Unjustified Absence), English as a Second Language (ESOL), Early Leaving exemptions, Home-schooling, Special Schooling, and many others.
MoH Cancer Registration Events Count	Numeric	Count of cancer registration events in the applicant household in the four years prior to application date. Used as indicative of urgent and serious medical need. Multiple instances of cancer for the same individual are treated as separate events, and counted separately.
MoH Chronic Conditions Registration Events Count	Numeric	Count of chronic conditions registered for applicants in the household in the four years prior to application date. Indicative of persistent medical need. Multiple instances of chronic conditions for the same individual are counted separately.
MoH General Medical Subsidy Claims Cost	Numeric	Total costs related to GMS claims made by the individuals in the applicant household in the four years prior to application, by 12 months periods. GMS claims have been steadily decreasing since 2005, and being replaced with the registration with Medical Practitioners, but high GMS claims may be indicative of a sub-group moving around from place to place without registration with a practitioner.
MoH Hospitalisation Cost	Numeric	Total costs related to hospital admission events, summed up for all members of the applicant household in the four years prior to application date, by 12 months periods.

Features	Type	Description
MoH Lab Test Event Cost	Numeric	Total costs related to laboratory tests undergone by the individuals of the household in the four years prior to application, by 12 months periods. Indicative of frequent medical issues and expenses.
MoH Pharmaceuticals Dispensation Cost	Numeric	Total costs related to pharmaceuticals dispensation events for all individuals of the applicant household in the four years prior to the application date, by 12 months periods.
Older Adults Count	Numeric	Percentage of adults above the age of 70 in the applicant household.
Older Children Count	Numeric	Percentage of children and young adults in the applicant household between the ages of 6 to 19 (inclusive).
Working Age Adults Count	Numeric	Percentage of adults in the applicant household above 19 and below 70.
Young Children Count	Numeric	Number of children aged five years and below in the applicant household
Primary Wage	Numeric	Total declared wage and salary for the primary applicant over the four years prior to application.
Primary Benefit	Numeric	Total received benefit payments for the primary applicant over the four years prior to application.
Primary Benefit to Wages Ratio	Numeric	Logarithm of the ratio of benefits to wages received over the four years prior to application

Table 12: Discarded variables description

Variable	Type	Description
Current Meshblock	Categorical	Current meshblock of the primary applicant. This may be useful in deriving the deprivation index using the meshblock decile values.
Current Territorial Authority Code	Categorical	This variable describes further information about the current location of the primary applicant. This variable is not used because of the number of levels in the categorical variable. It may be possible to identify useful groupings and collapse this category to a manageable number in the future.
Household Type	Categorical	This variable describes the type of household, in terms of composition and relationships. For the cohort of interest, a large majority of records have missing values for this attribute and cannot be imputed.
No Location Preference Flag	Categorical	This flag indicates whether the applicant specified a location preference. In case the applicant specified more than three preferred locations, they are treated as having no preference. Not having a location preference may be strongly correlated with high and pressing need for housing. However, the data quality is relatively poor, with a lot of missing values. It is not known if these are random missing values or if this is indicative of an underlying driver.
Preferred Location	Categorical	This variable describes whether the applicant specified a particular location of interest for housing in the application. If combined with the number of houses available in the stated location matching the applicant preference, this could be a strong predictor for the applicant getting housed or not. However, the data quality of the attribute is relatively poor for the time being and this seems to be a free text field. Extracting useful information out of this variable may require further analysis.

Appendix B The Social Investment Analytical Layer (SIAL)

Creating the SIAL

Need for SIAL

The IDI holds administrative data from government agencies and NGOs across New Zealand, which has been linked at the individual level and anonymised. Agencies, however, do not capture data in a consistent way, making it difficult and time-consuming to work with data from different agencies.

The SIAL is built from data in Statistics NZ's IDI. It arranges the data held into a standard format, making it easier and faster for analysts to understand and use.

All of SIU's current and future analysis uses the SIAL. Agencies other than SIU are experimenting with the SIAL. The SIU's objective is to share the code so other parties can use it to create the tables they need in the SIAL format.

Discussions are underway for Statistics NZ to eventually be responsible for the deployment of the code to a production environment. This includes tables already generated and saved in SIU's development environment (sandpit), and the regular production of the SIAL tables. This will make them available for all authorised IDI users to access. In particular, Statistics NZ will be responsible for the on-going maintenance of these tables to ensure they are updated with each quarterly IDI refresh.

SIAL standardised table format – current state

To date, the SIAL shows information from ACC, COR, CYF, MoE, MoH, Justice, MSD and NZ Police. This has been reformatted into events-structured tables, which make up the SIAL.

Table 13 displays the variables available in each table and gives an example using the IDI table for CYF substantiated abuse.

Table 13: Variables available in the events structured tables within SIAL

Variable	Format	Description	Example (CYF substantiated abuse table)
snz_uid	Numeric	Unique identifier	snz_uid value
Department	Character ($char3.)	Three letter abbreviation for the department the data is collected from	MSD
Datamart	Character	Three letter abbreviation for the datamart the data is collected from. If there is no datamart in the original table this is the area within a department the data comes from	CYF
Subject_area	Character	Three letter abbreviation for the subject area of the event. If this in not available in the original table, this variable is developed to describe the subject area of the table	ABE (CYF abuse finding events)
Start_date	Date (DATETIME)	Event start date	Date of abuse finding, e.g. 30JUN2014 00:00:00

Variable	Format	Description	Example (CYF substantiated abuse table)
End_date	Date (DATETIME)	Event end date	Date of abuse finding, e.g. 30JUN2014 00:00:00. Note: as abuse findings are point in time events, that if their duration length is 0 days, they share the same start and end dates
Revenue (optional)	Numeric	Revenue generated by event	NA. No revenue generated from CYF abuse finding events table. Note: to date only the IR table has revenue
Cost (optional)	Numeric	Lump sum ***direct*** costs of event for the event duration. Only available when costs for an event attributible to an individual are available	NA. No unit level costs available in the CYF abuse finding events table
Total_cost (optional)	Numeric	Lump sum ***direct*** costs of event for the event duration + lump sum ***indirect*** costs of event for the event duration. Note: indirect costs for agencies available in the IDI, and attributable to the individual-level are rare	
Event_type	Character	Three letter codes representing additional details about the event. These codes are mapped in classification tables	Abuse type (SEX, PHY, EMO, …)
Event_type_n (optional)	Character ($char3.)	Three letter codes representing additional details about the event. These codes are mapped in classification tables	Did not specify extra details for these events tables

Example of measures generated from the SIAL tables

The benefit of having data structured into events is to make it easier and faster to understand the experiences of an individual's lifetime. The user is able to create metrics within a specified date range. These metrics include, but are not limited to:

- Total cost of an event
- Total revenue generated by the event
- Total duration of an event
- Number of times an event has occurred
- Duration since the first event of the same type
- Duration since the last event of the same type.

This also allows the user to quickly build up a picture of a person's life. Figure 11 shows the life experiences of a fictitious individual (Sam) as a series of interactions with various government agencies throughout his life (referred to as events).

By using the SIAL, all of these metrics about education, health and welfare can be easily and quickly produced. Before SIAL, this would have been a labour-intensive and time-consuming task.

Standardised event tables were generated to allow a consistent set of analysis to be performed across every table. These events tables contain start and end dates, the type of event and the costs (where available). Table 14 contains an example of an event table.

Figure 11: Example of a person (Sam) experiencing various events during their life course

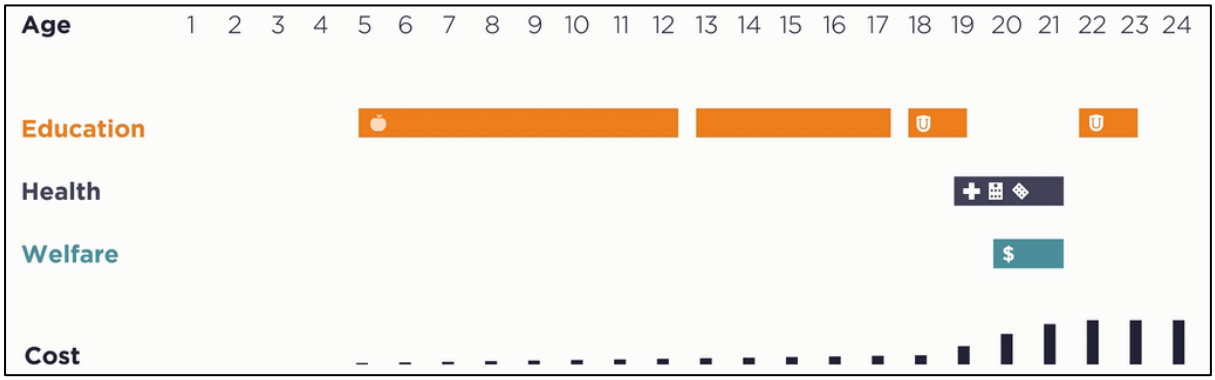


Table 14: Example of an events table using synthetic data

Snz_uid	Department	Datamart	Subject area	Start_date	End_date	Event type	Cost
1234	MSD	BEN	T1	01/01/2013	31/12/2014	JSS	$10,987.65
9876	MSD	BEN	T1	03/03/2014	12/12/2014	YPP	$1,234.56
1234	MSD	BEN	T1	30/06/2015	30/09/2015	JSS	$1,234.56

Table 15 contains a full list of SIAL table subject areas currently available and when data is available from.

Note: since the purpose of the SIAL is standardisation of tables, rather than applying business rules, all the IDI documentation for these tables can be used to identify data quality issues. For example, the data dictionary for health mentions potential duplicate rows and also lack of coverage in some datasets. Additional information about the data tables, including an online discussion board, is available at Meetadata – the IDI online forum[8] (which does not require IDI access). Access to the forum is available by contacting meetadata@stats.govt.nz.

Table 15: Current SIAL tables and date the data is available from

Agency	Subject area	Description	Available from
MSD	T1	Main benefits	1993
MSD	T2	Supplementary benefits	1993
MSD	T3	Hardship payments	1993
MoE	INT	Student interventions (e.g. truancy)	1997*
MoE	ENR	Primary and secondary school enrolments	2007
MoE	ECE	Early childhood education	2007*
MoE	TER	Tertiary enrolments	1994

[8] http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/idi-resources.aspx

Agency	Subject area	Description	Available from
IR	W&S	Wages and salary	1999
IR	PEN	Pension payments	1999
IR	CLM	ACC weekly compensation	1999
IR	STU	Student allowance	1999
IR	BEN	Benefit payments	1999
IR	PPL	Paid parental leave	2002
ACC	INJ	ACC injuries	1994
MoJ	CHA	Court charges	2004
COR	SAR	Corrections sentencing and remand	1970
MoH	CHR	Chronic conditions	1988
MoH	PFD	Publically funded hospital discharges	1989
MoH	CAN	Cancer registrations	1995
MoH	GMS	General medical subsidy (after hours and unenrolled patients)	2002
MoH	LAB	Lab tests	2003
MoH	PHA	Pharmaceutical dispensings	2005
MoH	NPA	National non-admitted patients (outpatients and emergency department)	2007
MoH	PRI	PRIMHD – mental health collection	2008*
MoH	B4S	Before school health checks	2011
HNZ	REG	HNZ register waitlist events	2000
CYF	CEC	CYF events	1992*
CYF	ABE	CYF abuse findings	1994*
POL	VIC	Police records of victimisations	2014
POL	OFF	Police records of offenders	2009

* Known coverage and data quality issues

In addition to the SIAL, SIU has created a tool – Social Investment Measurement Map (SIMM) – to list outcomes that can be measured in SIAL. This enables new authorised IDI users to see what measures can be derived from SIAL and IDI tables. It will be updated as feedback is received. The SIMM is available on the SIU's website[9].

The rollup process and use of history, profile, forecast windows

The information in the events tables can be aggregated into what is referred to as a rolled-up table. This produces aggregate level measures for an individual over a given timeframe. Rather than having a long events table with one row per event, there is now a wide rolled-up table with one row per ID (this could be an ID of an individual person, a household or some other unique ID).

Table 16 shows a sample set of columns for a rolled-up version Table 14.

The columns of the rolled up table contain measures of duration, counts and costs broken down by event type. For example, by using the MSD data there can be:

- A total duration on benefit
- A count of the benefit spells

[9] https://www.siu.govt.nz/tools-and-guides/measurement-map/

- How recent the benefit was
- The total cost of the benefit.

This can also be drilled down into benefit subtypes to produce what is known as a wide dataset.

Table 16: Example of a rolled up table

Snz_uid	f_msd_ben_t1_cnt	f_msd_ben_t1_dur	f_msd_ben_t1_cst
1234	**2**	**793**	**$12,222.21**
9876	**1**	**284**	**$1,234.56**

The naming convention for the columns comprises of the following acronyms:

<W>_<XXX>_<YYY>_<ZZZ>_<AAA>

- W = the window. This can be either the window prior to the intervention – the profile window (p) or the window after exposure to an intervention, i.e. the forecast window (f). The code used distinguishes between two other windows:
    - Analysis window (a) is the period following the forecast window
    - History window (h) that occurs before the profile. However, these aren't used most of the time.
- XXX = the department where the data comes from, e.g. MSD, MoH and so on. When data has come from combined set of departments the agency MIX tends to be used.
- YYY = the datamart (e.g. BEN for benefits).
- ZZZ = subject area (e.g. T1 for tier one main benefits).
- AAA = one of the following metrics: count (cnt), duration (dur), cost (cst), days since the first event (dsf), days since the last event (del).

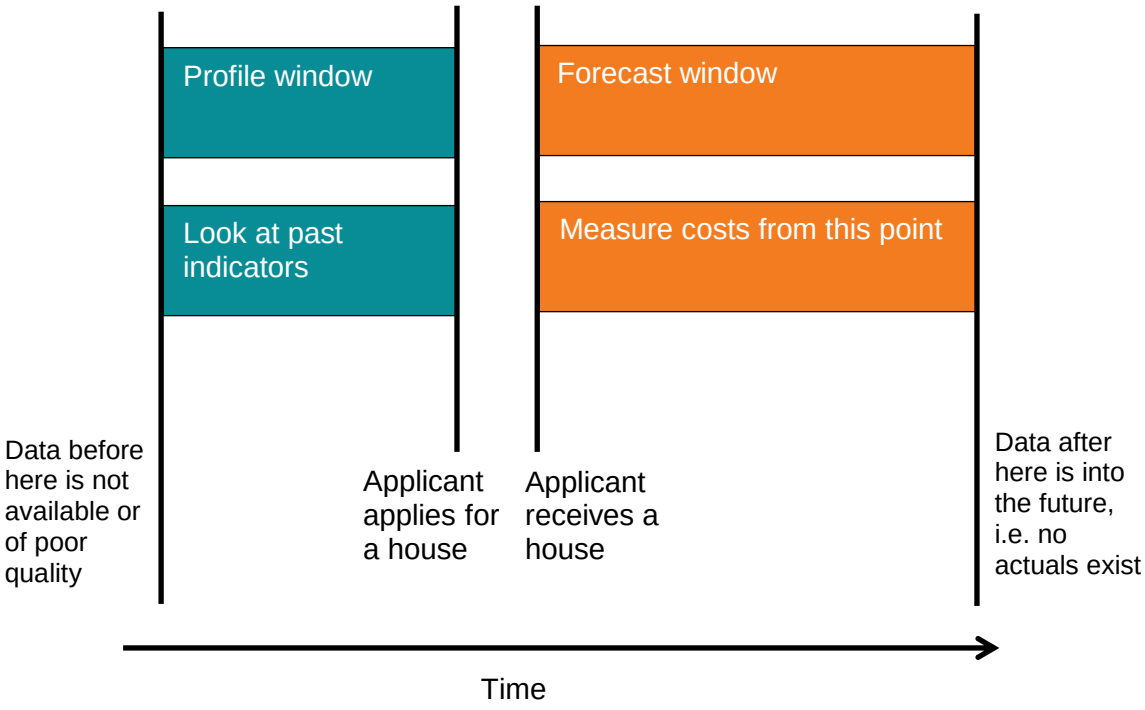
The events tables have good coverage between 2001 and 2014. Some events tables have longer coverage, some have less.

Example of how this process works for social housing

In the test case, it was defined that the intervention occurs from the date someone enters a social house. Costs and benefits were to be measured from this point onwards. This period is referred to in the modelling section as the forecast window. It is the window where the ROI from can be measured (see Figure 12).

For those who applied and were not placed in social housing, the intervention date was set to the date they applied for a house. The original reasoning for this was to measure the effects of social housing, so it was necessary to wait until clients were housed before starting to measure outcomes. Rolled-up measures for both the profile and forecast windows have been generated.

Figure 12: Description of the profile and forecast windows and how they relate to social housing



The following events tables have been rolled up during the profile window and feature count and duration measures aggregated at a high level, so there is a single measure for each person (who has interacted with the given agency), for each table:

- CYF client events
- CYF abuse
- ACC claims
- MSD (Tier 1 main benefits, tier 2 supplementary benefits and tier 3 hardship payments)
- MoE student interventions
- Corrections spells
- MoH mental health (PRIMHD)
- MoH cancer registrations
- MoH chronic conditions
- MoH medical subsidies (GMS)
- MoH ED, and outpatients (National Non-admitted patient collection (NNPAC))
- MoH lab tests
- MoH publically funded hospital discharges (PFD)
- MoH pharmaceuticals.

The following tables were rolled up for ROI calculations:

- CYF client event costs
- ACC non earner medical costs
- MSD tier 1 benefits
- MSD tier 2 benefits excluding working for families which is picked up on the IR side
- MSD tier 3 non recoverable costs

- Corrections costs
- Mental health costs (PRIMHD)
- Education costs
- Before school checks
- General Medical Subsidy costs
- NNPAC costs
- Lab costs
- Public hospitalisation costs
- Pharmaceutical costs
- IR revenue (W&S tax component)
- International Rescue Committee (IRC) costs (ACC weekly compensation claims, paid parental leave, student allowance and pension payments).

A description of the data and a data dictionary for the SIAL will be available separately.

Key lessons and future improvements from data preparation and creating standardised tables

One of the key goals for this test case was to see whether it was possible to create a standardised methodology in a reusable matter for future analysis. The SIAL achieved this goal and can be used for future analysis.

A couple of other agencies are currently testing the SIAL tables in their own analyses. The following has been highlighted to SIU's test team:

- The need for reusable structures cannot be emphasised enough. The SIU was able to quickly make changes to this test case post-feedback because of the standardised tables. Other groups who have tested the SIAL have been able to quickly build a lifetime view of a person, saving considerable amounts of time.
- Currently the report points to data dictionaries and online forums where data quality issues have been noted. One future improvement could be to create a section within the report noting quality issues with the underlying data that would save readers going to the external sources to locate the information. It would also make any future extensions of this work by other agencies or groups easier.
- When creating two groups for comparison, careful consideration should be given to over what time period to measure outcomes. Originally, SIU chose to measure outcomes from the time of housing for those housed and the application date for those who were not housed (so outcomes were not recorded while waiting for a social house). Feedback indicated this had the potential to introduce a time bias. The average time someone waits to be housed is around three months, so the effect of time bias should be minimal. In future, the SIU will probably measure outcomes for two groups over the same timeframe.

Appendix C Cohort descriptive statistics

By sentencing and remand (S&R) count

S&R	Housed	Other exit
0	8763	9366
1	876	846
2	378	348
3	213	219
4	138	126
>=5	264	291

By number of children in the household

Number of children in household (total)	Housed	Other exit
0	4461	5202
1	2196	2835
2	1914	1917
3	1224	750
>=4	840	492

By number of young children (under five) in the household

Young children in household	Housed	Other exit
0	6792	7512
1	2352	2586
2	1068	912
>=3	420	183

By sole earner indicator

Sole earner indicator	Housed	Other exit
No income	3141	2979
Primary is sole earner	5373	6303
Not sole earner	2118	1914

By application main reason

Application main reason	Housed	Other exit
BETTER UTIL	99	81
CUSTODY ACCS	54	48
DISCRIMINATN	30	48
EMP OPPORT	27	75
FAMILY REASN	1158	1464
FINANCIAL	1302	2187
FIRE DAMAGE	24	12
HEALTH	1185	1230
HNZ SERVICES	447	606
HOMELESSNESS	1170	744
HOME SOLD	147	144
HSE FOR SALE	171	258
INADEQUATE	585	567
MODIFICATION	123	48
NEIGHBOUR IS	36	78
OVERCROWDING	2553	2136
PERS SAFETY	303	297
SPECIAL NEED	183	189
TENANCY TERM	1032	984

By total score

Total score	Housed	Other exit
A	1146	399
B	7068	5049
C	1824	3972
D	600	1779

The ‘P_’ prefix denotes the profile window. Variables refer to the household unless otherwise indicated.

Variable name	Variable name (long)	Group	Mean	SD	Median	Mean Abs Dev.	Skewness
accs_score	Access score	Housed	1.61	0.72	1.00	0.00	0.83
		Other exit	1.40	0.60	1.00	0.00	1.26
adeq_score	Adequacy score	Housed	1.04	0.36	1.00	0.00	8.07
		Other exit	1.02	0.22	1.00	0.00	13.78
afford_score	Affordability score	Housed	1.50	0.81	1.00	0.00	1.55
		Other exit	1.55	0.82	1.00	0.00	1.40
suit_score	Suitability score	Housed	1.98	0.96	2.00	1.48	0.54
		Other exit	1.60	0.77	1.00	0.00	1.06
sustain_score	Sustainability score	Housed	2.70	0.77	3.00	0.00	-0.48
		Other exit	2.25	0.83	2.00	1.48	-0.08
Primary_age	Age of primary	Housed	38.49	14.74	36.00	14.83	0.79
		Other exit	37.60	14.75	35.00	14.83	0.81
primary_total_wage	Total W&S for primary applicant in last 48 months	Housed	2379.83	3937.37	376.87	558.75	2.19
		Other exit	2432.96	3915.55	517.24	766.85	2.30
applicant_total_benefit	MSD benefit receipt amount in the previous 48 months (primary applicant)	Housed	26967.62	18912.71	28491.81	25393.61	-0.04
		Other exit	25704.70	18384.13	26162.09	25886.05	0.06
ben_to_wage_log_ratio	Wage to benefit ratio in the past 48 months for main applicant (natural log)	Housed	4.53	4.71	3.99	4.62	-0.49
		Other exit	4.19	4.65	3.58	4.26	-0.39

Variable name	Variable name (long)	Group	Mean	SD	Median	Mean Abs Dev.	Skewness
P_ACC_CLA_INJ_Y1 _cost	Costs related to ACC claims (injuries) in last 12 months	Housed	152.67	3993.33	0.00	0.00	98.01
		Other exit	97.07	639.26	0.00	0.00	26.00
P_ACC_CLA_INJ_Y2 _cost	Costs related to ACC claims (injuries) in last 12 to 24 months	Housed	83.06	675.73	0.00	0.00	33.52
		Other exit	60.84	361.44	0.00	0.00	29.22
P_ACC_CLA_INJ_Y3 _cost	Costs related to ACC claims (injuries) in last 24 to 36 months	Housed	58.24	340.61	0.00	0.00	23.04
		Other exit	66.91	604.69	0.00	0.00	28.49
P_ACC_CLA_INJ_Y4 _cost	Costs related to ACC claims (injuries) in last 36 to 48 months	Housed	63.09	724.23	0.00	0.00	57.76
		Other exit	54.13	401.96	0.00	0.00	22.53
P_CYF_EVE_CNP_Y1 _cost	Costs of CYF CNP related events in last 12 months	Housed	199.84	2212.40	0.00	0.00	22.44
		Other exit	139.25	1634.94	0.00	0.00	18.94
P_CYF_EVE_CNP_Y2 _cost	Costs of CYF CNP related events in last 12 to 24 months	Housed	155.53	2073.86	0.00	0.00	26.92
		Other exit	116.96	1970.69	0.00	0.00	46.40
P_CYF_EVE_CNP_Y3 _cost	Costs of CYF CNP related events in last 24 to 36 months	Housed	122.53	1752.19	0.00	0.00	26.84
		Other exit	93.18	1307.04	0.00	0.00	22.87
P_CYF_EVE_CNP_Y4 _cost	Costs of CYF CNP related events in last 36 to 48 months	Housed	76.12	1127.82	0.00	0.00	31.19
		Other exit	74.41	1014.01	0.00	0.00	22.52

Variable name	Variable name (long)	Group	Mean	SD	Median	Mean Abs Dev.	Skewness
P_CYF_EVE_YJU_Y1 _cost	Costs of CYF YJU related events in last 12 months	Housed	7.21	289.74	0.00	0.00	51.51
		Other exit	5.50	317.55	0.00	0.00	86.27
P_CYF_EVE_YJU_Y2 _cost	Costs of CYF YJU related events in last 12 to 24 months	Housed	4.35	172.07	0.00	0.00	49.20
		Other exit	8.28	350.32	0.00	0.00	64.14

Variable name	Variable name (long)	Group	Mean	SD	Median	Mean Abs Dev.	Skewness
P_CYF_EVE_YJU_Y3 _cost	Costs of CYF YJU related events in last 24 to 36 months	Housed	2.73	124.57	0.00	0.00	70.37
		Other exit	38.48	3763.92	0.00	0.00	102.96
P_CYF_EVE_YJU_Y4 _cost	Costs of CYF YJU related events in last 36 to 48 months	Housed	2.49	112.59	0.00	0.00	70.95
		Other exit	7.21	289.74	0.00	0.00	51.51
P_DoC_MMP_SR_Y1 _cost	Costs of S&R related events in last 12 months	Housed	620.86	3789.24	0.00	0.00	8.92
		Other exit	811.24	4721.84	0.00	0.00	7.90
P_DoC_MMP_SR_Y2 _cost	Costs of S&R related events in last 12 to 24 months	Housed	574.42	3841.87	0.00	0.00	10.05
		Other exit	788.86	4814.55	0.00	0.00	7.95
P_DoC_MMP_SR_Y3 _cost	Costs of S&R related events in last 24 to 36 months	Housed	531.21	3609.36	0.00	0.00	9.61
		Other exit	663.14	4413.62	0.00	0.00	8.65
P_DoC_MMP_SR_Y4 _cost	Costs of S&R related events in in last 36 to 48 months	Housed	459.92	3301.66	0.00	0.00	9.90
		Other exit	556.41	4038.37	0.00	0.00	10.00

Variable name	Variable name (long)	Group	Mean	SD	Median	Mean Abs Dev.	Skewness
P_IRD_INC_BEN_Y1 _cost	Costs related to payment of tier 1 benefit in last 12 months	Housed	9487.26	6248.07	10745.81	5903.83	0.03
		Other exit	8975.83	6046.01	10019.44	6618.53	0.02
P_IRD_INC_BEN_Y2 _cost	Costs related to payment of tier 1 benefit in last 12 to 24 months	Housed	7998.96	6219.35	9113.82	7264.09	0.19
		Other exit	7448.47	5988.02	8350.50	7683.15	0.22
P_IRD_INC_BEN_Y3 _cost	Costs related to payment of tier 1 benefit in last 24 to 36 months	Housed	7056.97	5881.78	7797.83	7751.92	0.30
		Other exit	6494.28	5576.68	6994.55	8037.57	0.29
P_IRD_INC_BEN_Y4	Costs related to payment of tier 1	Housed	6284.76	5501.25	6653.43	7809.51	0.37

Variable name	Variable name (long)	Group	Mean	SD	Median	Mean Abs Dev.	Skewness
_cost	benefit in last 36 to 48 months	Other exit	5719.85	5214.08	5954.00	8177.66	0.39
P_IRD_INC_W_S_Y1 _cost	Total declared W&S (household) in last 12 months	Housed	806.77	1481.54	8.29	12.29	2.55
		Other exit	800.14	1457.00	22.70	33.65	2.60
P_IRD_INC_W_S_Y2 _cost	Total declared W&S (household) in last 12 to 24 months	Housed	853.21	1506.84	17.65	26.17	2.31
		Other exit	843.26	1478.65	31.82	47.18	2.42
P_IRD_INC_W_S_Y3 _cost	Total declared W&S (household) in last 24 to 36 months	Housed	780.76	1390.17	11.57	17.15	2.34
		Other exit	755.61	1342.16	28.57	42.36	2.55
P_IRD_INC_W_S_Y4 _cost	Total declared W&S (household) in last 36 to 48 months	Housed	683.57	1262.07	0.00	0.00	2.48
		Other exit	668.45	1221.18	8.10	12.00	2.51

Variable name	Variable name (long)	Group	Mean	SD	Median	Mean Abs Dev.	Skewness
P_MOE_ENR_ENR_Y1_cost	Costs related to school enrolment in last 12 months	Housed	951.90	2390.40	0.00	0.00	3.28
		Other exit	658.56	1898.88	0.00	0.00	3.67
P_MOE_ENR_ENR_Y2_cost	Costs related to school enrolment in last 12 to 24 months	Housed	427.70	1411.59	0.00	0.00	4.19
		Other exit	304.97	1167.90	0.00	0.00	4.76
P_MOE_ENR_ENR_Y3_cost	Costs related to school enrolment in last 24 to 36 months	Housed	188.60	825.99	0.00	0.00	5.33
		Other exit	145.05	731.30	0.00	0.00	6.18
P_MOE_ENR_ENR_Y4_cost	Costs related to school enrolment in last 36 to 48 months	Housed	77.63	459.08	0.00	0.00	6.74
		Other exit	51.07	374.33	0.00	0.00	9.38
P_MOE_MOE_INT_cnt	Count of student interventions last 48 months	Housed	0.29	0.97	0.00	0.00	6.19
		Other exit	0.16	0.65	0.00	0.00	6.26

Variable name	Variable name (long)	Group	Mean	SD	Median	Mean Abs Dev.	Skewness
P_MOH_CAN_REG_cnt	Count of cancer registrations	Housed	0.02	0.13	0.00	0.00	7.53
		Other exit	0.01	0.11	0.00	0.00	9.80
P_MOH_TKR_CCC_cnt	Count of chronic condition events	Housed	0.19	0.51	0.00	0.00	3.20
		Other exit	0.16	0.47	0.00	0.00	3.88
P_MSD_CYF_ABE_cnt	Count of CYF events	Housed	0.60	2.03	0.00	0.00	6.57
		Other exit	0.39	1.66	0.00	0.00	11.51

Variable name	Variable name (long)	Group	Mean	SD	Median	Mean Abs Dev.	Skewness
P_MOH_GMS_GMS_Y1_cost	Costs related to GMS events in last 12 months	Housed	40.29	88.44	0.00	0.00	4.70
		Other exit	37.32	80.34	0.00	0.00	5.03
P_MOH_GMS_GMS_Y2_cost	Costs related to GMS events in last 12 to 24 months	Housed	44.14	94.36	0.00	0.00	4.18
		Other exit	42.72	95.45	0.00	0.00	5.07
P_MOH_GMS_GMS_Y3_cost	Costs related to GMS events in last 24 to 36 months	Housed	54.38	113.51	11.51	17.07	4.36
		Other exit	48.73	105.04	11.51	17.07	4.51
P_MOH_GMS_GMS_Y4_cost	Costs related to GMS events in last 36 to 48 months	Housed	41.11	101.69	0.00	0.00	5.03
		Other exit	28.81	78.56	0.00	0.00	5.67
P_MOH_LAB_TESTS_Y1_cost	Costs related to LAB TEST events in last 12 months	Housed	82.91	128.84	37.75	55.97	4.19
		Other exit	83.51	123.56	43.98	65.21	3.91
P_MOH_LAB_TESTS_Y2_cost	Costs related to LAB TEST events in last 12 to 24 months	Housed	71.63	119.08	25.72	38.14	6.35
		Other exit	66.89	108.42	21.41	31.75	3.84
P_MOH_LAB_TESTS_Y3_cost	Costs related to LAB TEST events in last 24 to 36 months	Housed	48.52	92.73	0.00	0.00	5.91
		Other exit	41.88	80.45	0.00	0.00	4.15
P_MOH_LAB_TESTS_Y4_cost	Costs related to LAB TEST events in last 36 to 48 months	Housed	19.31	61.39	0.00	0.00	9.20
		Other exit	12.55	41.25	0.00	0.00	5.54

Variable name	Variable name (long)	Group	Mean	SD	Median	Mean Abs Dev.	Skewness
P_MSD_BEN_AS_Y1 _cost	Costs related to AS payments in last 12 months	Housed	1663.54	1898.35	1028.98	1525.57	1.38
		Other exit	1752.14	1911.36	1132.24	1678.67	1.27
P_MSD_BEN_AS_Y2 _cost	Costs related to AS payments in last 12 to 24 months	Housed	1342.91	1699.47	629.87	933.84	1.49
		Other exit	1320.03	1646.42	654.23	969.96	1.48
P_MSD_BEN_AS_Y3 _cost	Costs related to AS payments in last 24 to 36 months	Housed	1094.04	1480.84	395.39	586.21	1.63
		Other exit	1083.58	1441.21	417.36	618.77	1.62
P_MSD_BEN_AS_Y4 _cost	Costs related to AS payments in last 36 to 48 months	Housed	971.34	1363.63	277.84	411.92	1.68
		Other exit	957.14	1330.24	287.71	426.56	1.63
P_MSD_BEN_T2_Y1 _cost	Costs related to tier 2 benefit payments in last 12 months	Housed	903.93	1726.76	28.12	41.69	3.49
		Other exit	833.09	1626.02	21.37	31.69	4.06
P_MSD_BEN_T2_Y2 _cost	Costs related to tier 2 benefit payments in last 12 to 24 months	Housed	756.03	1527.11	0.00	0.00	3.47
		Other exit	670.75	1425.20	0.00	0.00	4.07
P_MSD_BEN_T2_Y3 _cost	Costs related to tier 2 benefit payments in last 24 to 36 months	Housed	549.08	1247.18	0.00	0.00	3.73
		Other exit	467.83	1126.03	0.00	0.00	4.61
P_MSD_BEN_T2_Y4 _cost	Costs related to tier 2 benefit payments in last 36 to 48 months	Housed	396.98	1036.28	0.00	0.00	4.62
		Other exit	338.21	957.82	0.00	0.00	6.30

Variable name	Variable name (long)	Group	Mean	SD	Median	Mean Abs Dev.	Skewness
P_MSD_BEN_T3_Y1 _cost	Costs related to tier 3 benefit payments in last 12 months	Housed	264.92	488.25	114.29	169.44	4.79
		Other exit	241.92	498.56	95.24	141.20	5.41
P_MSD_BEN_T3_Y2 _cost	Costs related to tier 3 benefit payments in last 12 to 24 months	Housed	214.96	465.32	45.35	67.24	4.82
		Other exit	200.81	456.38	0.00	0.00	4.96
P_MSD_BEN_T3_Y3 _cost	Costs related to tier 3 benefit payments in last 24 to 36 months	Housed	190.41	417.21	0.00	0.00	4.53
		Other exit	174.89	421.03	0.00	0.00	5.41
P_MSD_BEN_T3_Y4 _cost	Costs related to tier 3 benefit payments in last 36 to 48 months	Housed	165.21	378.70	0.00	0.00	5.08
		Other exit	144.17	359.91	0.00	0.00	6.08
P_old_adult	Number of adults over 65 in application	Housed	0.10	0.34	0.00	0.00	3.63
		Other exit	0.09	0.32	0.00	0.00	3.93
P_wk_age_adult	Number of working age adults in application	Housed	1.08	0.55	1.00	0.00	0.63
		Other exit	1.03	0.53	1.00	0.00	0.53
P_older_child	Number of children over five in application	Housed	0.73	1.09	0.00	0.00	1.73
		Other exit	0.55	0.93	0.00	0.00	2.08
P_young_child	Number of children under five in application	Housed	0.55	0.85	0.00	0.00	1.60
		Other exit	0.44	0.72	0.00	0.00	1.63

Appendix D Variable transformation rules

This section allows readers to see the exact rules used to construct the transformed variables.

Table 17: Variable transformation rules

Features	Transformation Rules
12-month Household Income	Natural Log Transformation $$log\ income = \log_e(1 + 12month\_household\_income)$$
Age Category	Binning of primary applicant's age *if age between (0,19) then '(0-19)'* *else if age between (19, 35) then '(19, 35)'* *else if age (35, 70) then '(35, 70)'* *else '(>70)'*
Prioritized Ethnicity	Collapse variable levels due for classes with low counts *if primary_ethnic_ind = 'O' then primary_ethnic_ind = 'Z';*
Wage to Benefit Ratio	Natural Log Transformation $$ln\_prim\_wg\_ben\_ratio\_12mnth = \log_e(1 + primary\_wage\_12\_mnth) - \log_e(1 + primary\_benefit\_12\_mnth)$$

Appendix E Gradient-boosting model: variable importance

Table 20 gives an indication of what variables are useful in predicting whether or not a household will receive social housing. As expected, the SAS scores have the highest gains.

Table 20: Ranked feature importance table (only the top 15 variables are shown)

Rank	Feature	Gain	Cover	Cumulative gain
1	hnz_na_analy_score_sustain_text	0.253	0.095	0.253
2	hnz_na_analy_score_suitably_text	0.116	0.061	0.370
3	hnz_na_bedroom_required_cnt_nbr	0.052	0.037	0.422
4	hnz_na_analysis_total_score_textB	0.039	0.007	0.461
5	hnz_na_analy_score_access_text	0.039	0.041	0.501
6	P_MSD_BEN_AS_Y1_cost	0.037	0.059	0.538
7	hnz_na_analysis_total_score_textC	0.028	0.019	0.566
8	P_IRD_INC_BEN_Y1_cost	0.026	0.034	0.592
9	hnz_na_hshd_size_nbr	0.023	0.028	0.615
10	ben_to_wage_log_ratio	0.018	0.018	0.633
11	P_IRD_INC_BEN_Y4_cost	0.015	0.019	0.648
12	primary_total_ben	0.014	0.016	0.661
13	P_IRD_INC_BEN_Y2_cost	0.012	0.018	0.674
14	age_band50-64	0.012	0.028	0.685
15	hnz_na_analysis_total_score_text_A	0.011	0.009	0.696
16	hnz_na_analysis_total_score_textD	0.011	0.004	0.707
17	P_MSD_BEN_AS_Y4_cost	0.010	0.012	0.717
18	P_old_adult	0.010	0.010	0.727
19	P_MOE_ENR_ENR_Y1_cost	0.010	0.015	0.736
20	P_MOH_PFH_PFHD_Y2_cost	0.010	0.014	0.746

Appendix F Tuning for the gradient-boosting model

Gradient boosting models require tuning to determine the best set of parameters to use.

Table 18 shows the depth, step-size shrinkage, number of trees, row and column sub-sampling parameters, along with measures of how well the model with the given parameters performed. The shaded row indicates the best configuration.

Table 18: XGBoost model parameter selection and tuning

Model no.	Depth	Step-size shrinkage	Row subsampling	Column subsampling	No. of trees with early stopping	AUC for ROC	Classification error
1	3	0.1	0.75	0.6	137	0.749	0.319
2	5	0.1	0.75	0.6	98	0.747	0.318
3	7	0.1	0.75	0.6	74	0.742	0.323
4	3	0.01	0.75	0.6	96	0.721	0.343
5	5	0.01	0.75	0.6	41	0.729	0.334
6	7	0.01	0.75	0.6	66	0.735	0.327
7	3	0.001	0.75	0.6	29	0.716	0.346
8	5	0.001	0.75	0.6	41	0.726	0.336
9	7	0.001	0.75	0.6	63	0.732	0.334
10	3	0.1	0.9	0.6	183	0.751	0.316
11	5	0.1	0.9	0.6	61	0.745	0.323
12	7	0.1	0.9	0.6	92	0.742	0.324
13	3	0.01	0.9	0.6	24	0.715	0.346
14	5	0.01	0.9	0.6	184	0.735	0.328
15	7	0.01	0.9	0.6	129	0.738	0.328
16	3	0.001	0.9	0.6	46	0.716	0.345
17	5	0.001	0.9	0.6	48	0.726	0.338
18	7	0.001	0.9	0.6	53	0.732	0.333
19	3	0.1	0.75	0.8	98	0.748	0.318
20	5	0.1	0.75	0.8	144	0.748	0.318
21	7	0.1	0.75	0.8	82	0.741	0.325
22	3	0.01	0.75	0.8	29	0.715	0.344
23	5	0.01	0.75	0.8	82	0.730	0.334
24	7	0.01	0.75	0.8	96	0.734	0.328
25	3	0.001	0.75	0.8	36	0.714	0.345
26	5	0.001	0.75	0.8	36	0.724	0.336
27	7	0.001	0.75	0.8	44	0.730	0.333
28	3	0.1	0.9	0.8	120	0.749	0.315
29	5	0.1	0.9	0.8	73	0.746	0.322
30	7	0.1	0.9	0.8	81	0.747	0.318
31	3	0.01	0.9	0.8	63	0.717	0.344

Model no.	Depth	Step-size shrinkage	Row subsampling	Column subsampling	No. of trees with early stopping	AUC for ROC	Classification error
32	5	0.01	0.9	0.8	160	0.733	0.329
33	7	0.01	0.9	0.8	102	0.734	0.329
34	3	0.001	0.9	0.8	47	0.714	0.345
35	5	0.001	0.9	0.8	81	0.724	0.336
36	7	0.001	0.9	0.8	84	0.728	0.335

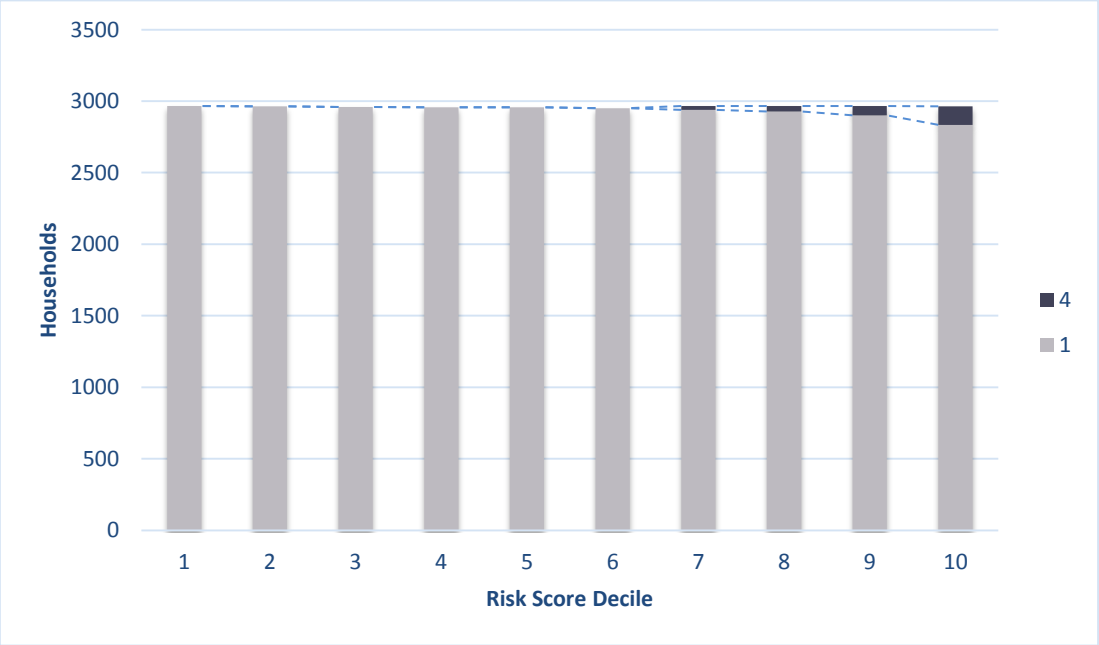
Appendix G Selected covariates by risk decile

Adequacy score

The Adequacy Score (Figure 13) has only two levels – 1 and 4 – for the cohort being modelled, where '4' is 'associated with a higher need for social housing.

Consequently, the model shows the same behaviour, by associating those households with a larger adequacy score with higher probability of getting housed. Sample sizes of households with an adequacy score of 4 are small – this is why the variable is rated low on the feature importance rank list.

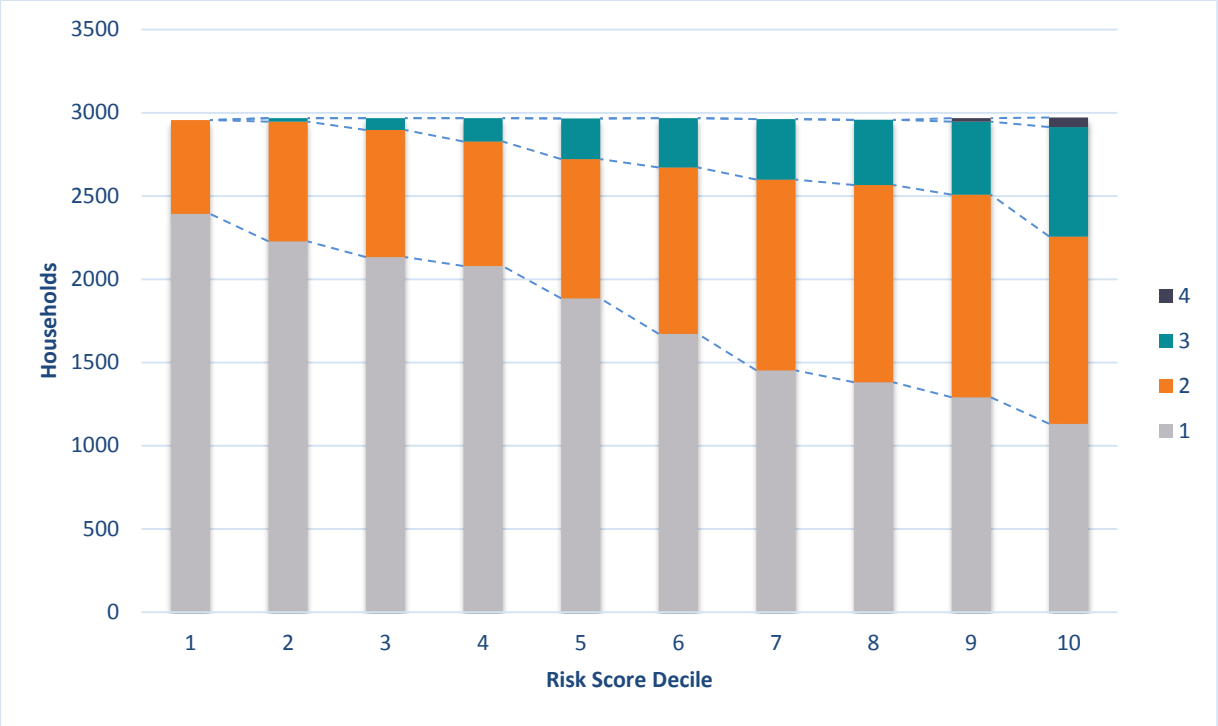
Figure 13: Adequacy score by propensity score decile



Accessibility score

The Accessibility score (Figure 14) is within the top 10 most important features for the model. The variable displays a similar behaviour to the other HNZ scores, in that there is a clear relationship between the score value and the probability of receiving social housing. The model reflects the same relationship, i.e. higher scores getting higher predicted probability for social housing.

Figure 14: Accessibility score by propensity score decile

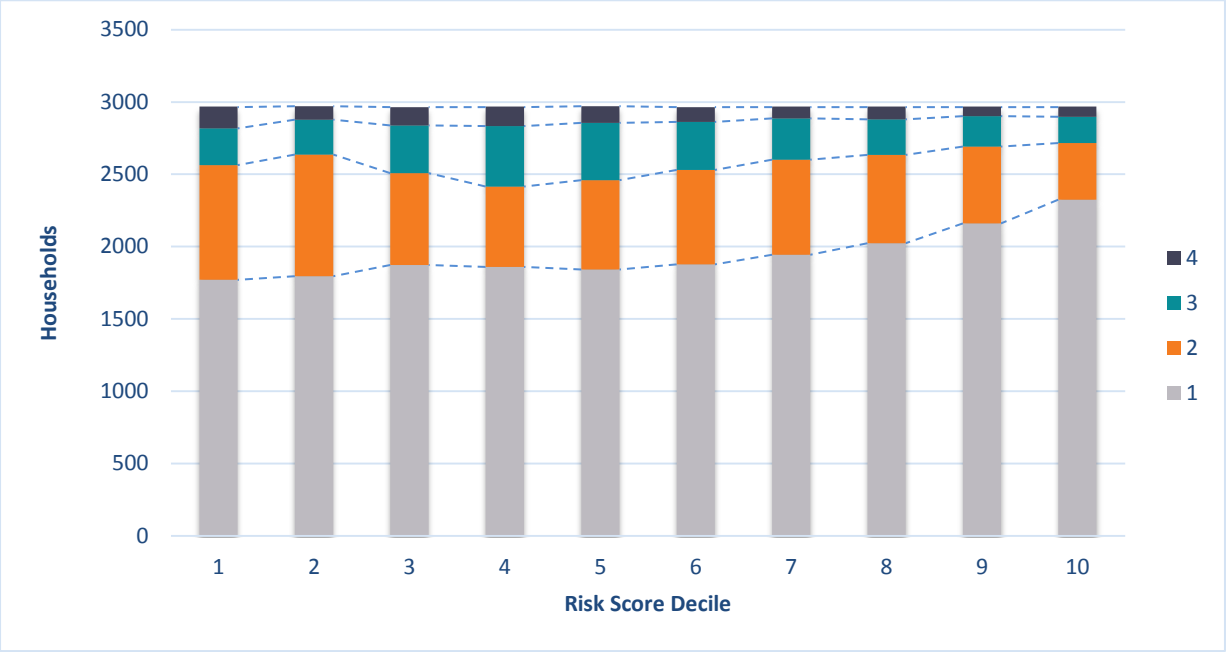


Affordability score

Affordability score (Figure 15) is rated quite low on the feature importance list for the model, and displays an uncertain relationship with the probability scores for obtaining social housing. The probability appears to increase with an affordability score of 1.

Households with a score of 3 peak towards the mid-decile range. This may be indicative of the variable's interaction with other variables used within the model, like region-level differences in affordability of private housing, or income levels. Therefore the behaviour of this variable in the model cannot be easily interpreted.

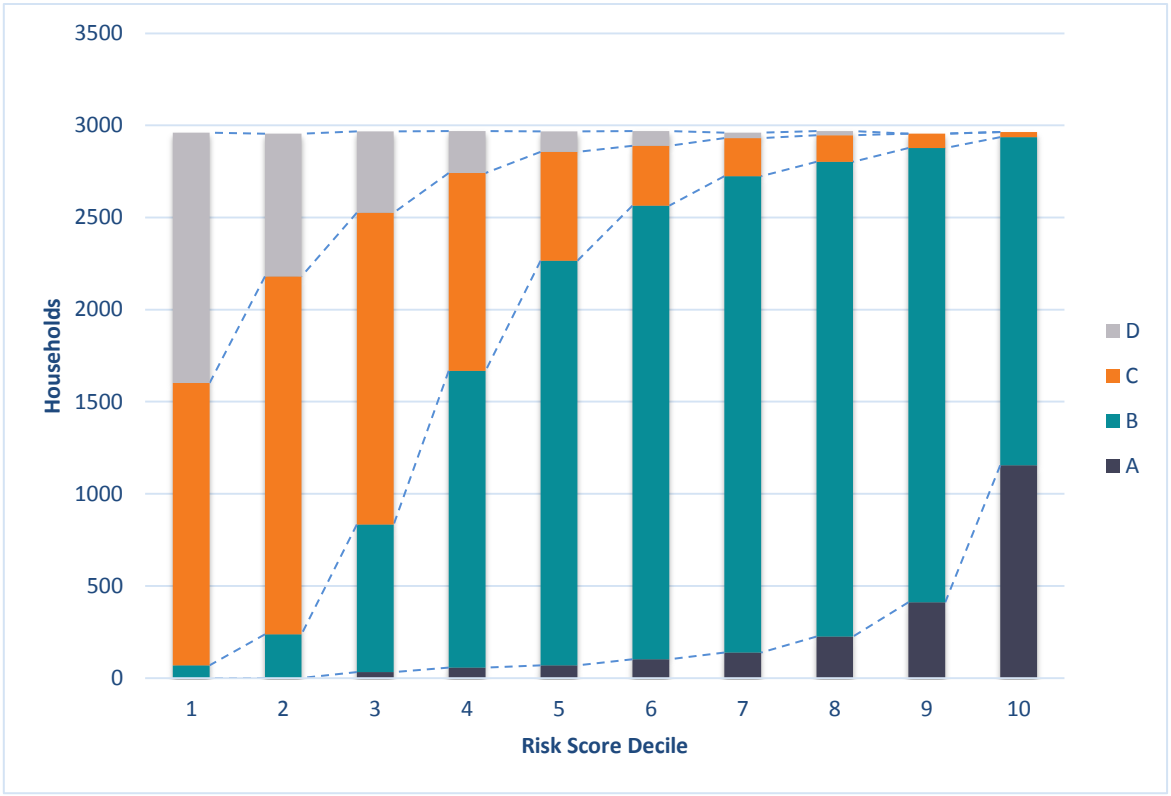
Figure 15: Affordability score by propensity score decile



Total score

The Total score (Figure 16) is the third most important variable in the feature importance list. A score of ‘A’ and ‘B’ are associated with a very high probability of obtaining social housing, and this relationship is clear from the model output below.

Figure 16: Total score by propensity score decile

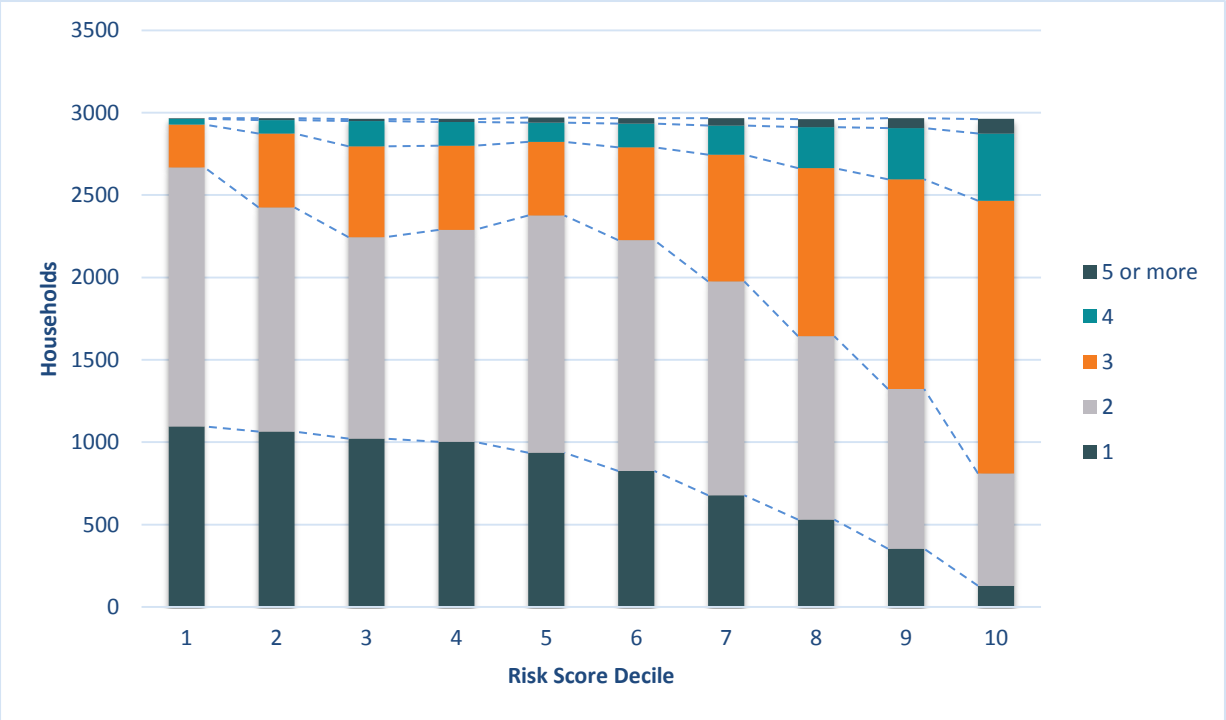


Bedrooms required

Bedrooms required (Figure 17) by the household tend to be closely correlated with the household size, and can be expected to show a similar relationship with the probability of obtaining social housing. An increase in the number of bedrooms required would signify a larger household size, and consequently a larger probability of being housed (see section on household size).

However, there could be differences based on household composition – a household with younger children may require fewer bedrooms compared to one with older children, even when the household sizes are comparable. The model output shows the expected relationship.

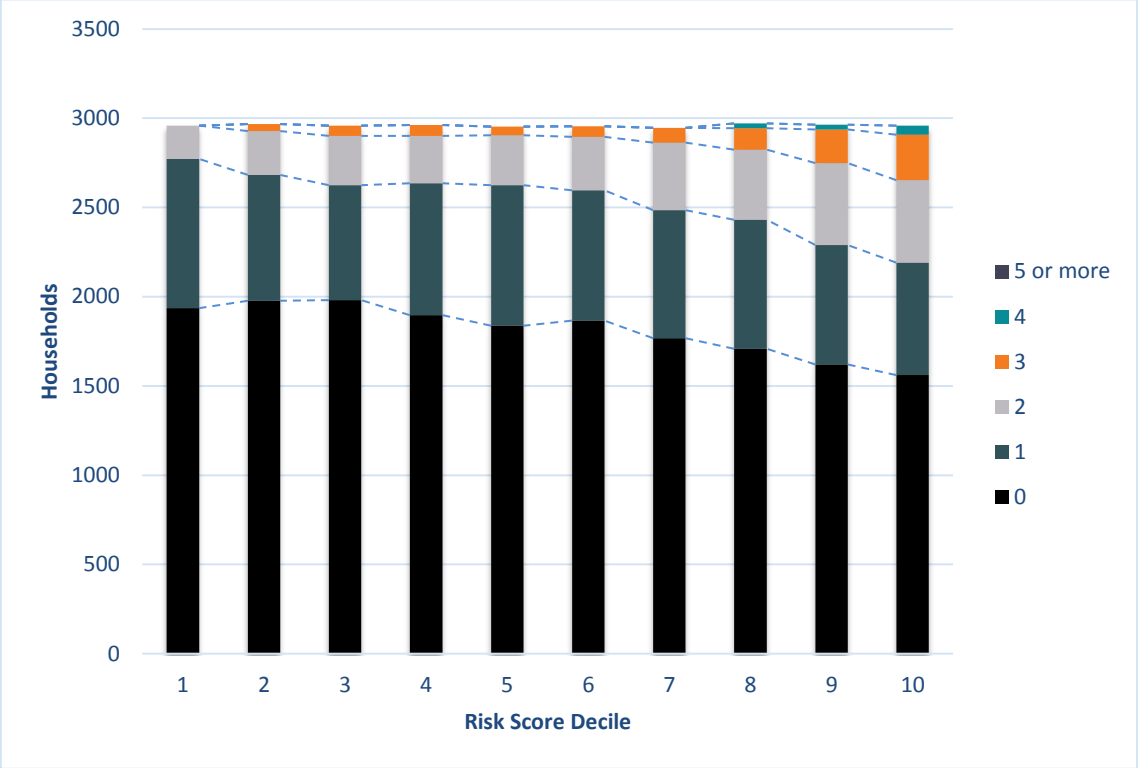
Figure 17: Bedrooms required by propensity score decile



Count of children under age five

As expected, the model displays an increase in the probability of obtaining social housing with an increase in the number of young children in the household (Figure 18). Here, young children are defined as those who are under the age of five.

Figure 18: Count of children under five years by propensity score decile



Appendix H More details on deriving the investment component

The derivation of the investment component for social housing is a complicated task. The IRRS payments form the investment but the values in the HNZ tables are inconsistent.

One would expect that IRR + IRRS = market rent. It is understood the market rent comes from a different source and does not reconcile IRR + IRRS.

To avoid this problem, a spreadsheet was provided from MSD's Social Housing Policy Group, containing information on the region bedroom size and IRRS + capital. This was transformed by SIU into Table 19.

Table 19: Table structure for policy cost table

TA Group	Bedroom number	Month 1 (t)	Month 2 (t_{+1})	Month n (t_{+n})
Auckland	1	$$	$$	$$
Auckland	2	$$	$$	$$
Auckland	3	$$	$$	$$
Auckland	4	$$	$$	$$
Auckland	5+	$$	$$	$$
Wellington	1	$$	$$	$$
…	…	…	…	…

When the IRRS + capital was incorporated there were some results that still didn't make sense, such as households paying 100% or more of the market rent, or households that essentially pay no market rent.

It is likely there is more variation than the high level regions and bedroom size takes into account. This variation could be caused by season variation, the business cycles or some other factor. To provide some more reasonable figures, a spline of the provided market rent + capital cost amounts was used. Specifically, a GAM spline was used for each bedroom number-region grouping with Generalised Cross Validation (GCV) used to determine an appropriate level of dampening.

Method for deriving the market rent + capital cost variable

The full method for constructing the market rent + cost of capital, including variable and table names, is outlined below:

1. ETL/tidy the adhoc_clean_tenancy_snapshot table from the IDI Sandpit (primary keys: legacy household UID, household UID, snapshot date).
2. Merge table with deduplicated/tidy version of houses_snapshot table from IDI Clean.
    a. Create 'house' and 'tenancy' flags
    b. Use bedroom number from houses_snapshot (set to 3 if 0 or missing; capped at five bedrooms).
3. Some more tidying to create a table (adhoc_clean_tenancy_house) with one row per house and month (like tenancy snapshot) with room number attached.
4. Use SAS proc format to map the TA from the HNZ tables to the TAs/region groups provided by policy (and attach these new region variables to the adhoc_clean_tenancy_house dataset).

5. Create table mr_mean (market rent mean) using adhoc_clean_tenancy_house where tenancy = 1 and bedroom number between 1 and 5 and year >= 2005.
    a. This table creates mean versions of market rent (for use later), IRR, IRRS, and weekly income.
    b. This table outputs one row per TA group per bedroom number and month.
6. Create table mr_cc (market rent + capital cost) that reads in the policy table and shapes it in the same way as mr_mean.
7. Merge mr_mean with mr_cc_3 to create mr_cc_updated
    a. This table includes the mr_cc variable (as well as the HNZ versions of MR, IRRS, etc.)
8. Use SAS proc tspline to create a splined version of the mr_cc values. The model is mr_cc ~ time (by TA group and bedrooms number).
    a. Create a ratio adjustment variable (mr_cc_adj) by dividing the spline value by the mean market rent for each group
    b. Create mr_cc_adj_fixed for rows with missing mr_cc values by taking an average by household number and date
    c. Add mr_cc, mr_cc_adj and mr_cc_adj_fixed to adhoc_clean_tenancy_house
9. Create final table – tenancy_housed_irrs
    a. Fill missing mr_cc_adj with mr_cc_adj_fixed
    b. Calculate irrs_adj variable
    c. Set IRRS values outside of $0 and $2,100 to zero.
    d. Calculate irrs_daily as irrs_adj / 7
    e. Final output is one row per house per month with irrs_adj cost attached.

Investment equation

Further details about the IRRS adjusted value explained in Step 9 above are outlined below:

$$irrs_{adj} = \max(abs\left(market\ rent \times\ mr_{cc_{adj}}\right) - abs(IRR), 0)$$

Where $market\ rent$ is the recorded market rent in the HNZ tenancy snapshot table for each house and month; $mr_{cc_{adj}}$ is the ratio adjustment between the splined 'market rent + capital cost' and the average market rent by region and bedroom number; and IRR is the recorded IRR in the HNZ tenancy snapshot table for each house and month.

The absolute values of these amounts are used because under HNZ's rental system the IRRS amounts were captured as positive values, while after migrating to a new system, they were recorded as negative values.

Notes:

- 4.2% of records are without an address, StatisticsNZ Unique Identifier (UID) or territorial authority (TA) variable.
- 0.004% of records are without a sensible weekly IRRS amount (and so were set to 0).
- From January 2005 to August 2015.
- No discount or CPI applied.

Joining tenancy_housed_irrs table with the hnz_social_hse_spells table

Due to lack of linking IDs joining, Table 19 is not a straightforward process. A workaround is outlined below:

> The tenancy_housed_irrs table did not contain an application UID field (the primary identifier for the chosen cohort). It was necessary to find a way to link the identifiers that it did have (house_uid and snapshot date), to the application UID on the cohort table.
>
> To do this, a (long) monthly version of the hnz_social_hse_spells table was created (which is in spell format) and (inner) joined it with the tenancy_housed_irrs table on house_uid and date – to create an intermediate table). It was then aggregated by application IDs and to get application spell totals for IRRS and 'IRRS + CC'.

Appendix I Decision log

This decision log (Table 20) is provided as an insight into the way SIU worked during the test case. It is not an exhaustive list of decisions made but decisions relating to the data affecting the analysis, which all analysts needed to be aware of. This decision log was added to as decisions were made throughout the test case. Names have been removed.

Table 20: Detailed decision log

Date	Decision	Description	Implications	Could this be improved for the next iteration? If so, provide details
10/06/2016	To define cohort of interest	Those who have applied will be used as our cohort. Housed will be cases, not housed will be controls.	Application data will be available for everyone in our cohort, and therefore a stronger propensity match can be created to understand the housing need. However, this means the propensity match won't consider everyone in NZ, so it won't be possible to see what the need for social housing is for everyone against those who are housed. It also means the chosen method will not be entirely reusable, as other interventions won't have HNZ data.	Not for social housing, but to make the method and code reusable the propensity match should be done for the whole NZ population, with no HNZ data.
20/06/2016	To decide profile and forecast windows	This has been documented in more detail as part of our technical documentation.	This sets out how many years of data contributes to the chosen propensity match and how many years of data contributes to the chosen ROI.	No.
20/06/2016	To have a cut off of two years after the application date	Only keep records with an end date in their application spell within two years of the start date for the same application spell.	The reason this is done is because in some cases a person is not housed or does not leave the register for three, four or even five years after application date. Even if these records were kept, a long run ROI would not be calculated until 2015.	No. Ideally you would keep everyone but there isn't enough data to make this feasible.

Date	Decision	Description	Implications	Could this be improved for the next iteration? If so, provide details
21/06/2016	To derive costs where they are not available in the IDI	This has been documented in more detail as part of our technical documentation.	Derived costs won't always be accurate, but they are good estimations given the information available. Estimating the costs is better than leaving it off and not being able to account for a large portion of social vote.	Yes. If cost data from agencies was available in the IDI then it would not have to be estimated and (possibly) it will be more accurate.
1/08/2016	Not to incorporate iMSD's (Forecasting &Costing) peer-review comments on the events tables	The peer-review began after the analysis on the tables, so there was no time to incorporate the comments into the analysis.	There are no major changes, but small quality checks. This is low risk for the analysis, but peer-review changes should be made in the next iteration.	Yes. Comments by iMSD will be documented and changes should be made to the tables before the next iteration.
1/08/2016	To use a Gradient Boosted Tree (GBT) model for the propensity scores calculation	This has been documented in more detail as part of the technical documentation.		No, not based of this decision to use this model, but the model itself could have more tests completed for the next iteration.
5/08/2016	To keep new applications and transfers in the 2005/06 cohort. This decision was overturned at a later date (11/08/2016).	Of approximately 40,000 applications in 2005/06 who were housed within two years, approximately 4,000 are transfer applications (10%). New applications and transfer applications will be used.	When calculating ROI over six years from being housed, it is cleaner to only include new applications, as transfer applications will have had prior exposure to the intervention (social housing). This means their outcomes may have already started changing as a result of being housed, before 2005/06. However, the policy question is what is the return of social housing as a process, therefore the investment bottom line is all those who are housed, new applications and transfers. Therefore it has been decided to keep all applications so all of the investment on those housed within two years from any application type in 2005/06 is captured. By producing the analytical output of ROI by duration of tenure in a social house, there is the	No.

Date	Decision	Description	Implications	Could this be improved for the next iteration? If so, provide details
			need to segment new applications and transfers, and to only do this part by new applications. Therefore this analysis will be based on 90% of the total cohort.	
8/08/2016	To recalculate the investment amount from numbers provided by policy	Policy will provided market rents + weekly maxima (capital) by region and bedroom number. This needs to be attached to the IDI data.	This is more work, as it was decided to not use the market rents amount in the IDI. The numbers provided by policy will also be averages by region and bedroom number, so won't be accurate by house (numbers in the IDI are by house). However, the value used will account for market rent + weekly maxima, which is more important to policy.	Yes, better data in the IDI at the house level that includes all investment amounts would make this more accurate.
8/08/2016	To impute for region and bedroom number where it can't be populated from the IDI	Region and bedroom number are needed to join the investment amounts from policy to our data. In cases where these cannot be populated they will need to be imputed.	Work is not finished but needs to be documented here.	Yes, better quality data in the IDI.

Date	Decision	Description	Implications	Could this be improved for the next iteration? If so, provide details
9/08/2016	Duplicates in the HNZ applications data	There are duplicates in the HNZ data that are problematic for our analysis. There are four different scenarios that create duplicates, which all need to be resolved, but the fourth is most problematic: 1. Multiple new applications in 2005/06 that are all housed 2. Multiple new applications in 2005/06 that are all not housed 3. A mixture of new and transfer applications in 2005/06 (a person requests a transfer after being housed) 4. Multiple applications in 2005/06 that are housed and not housed.	The solution to 1, 2, and 3 is to keep the record with the maximum start date within 2005/06. This means (within 2005/06) the first exposure to social housing, or a declined application, will be captured. Note: If a person also applied before 2005 they will not be captured in their first exposure, because it was decided to not look at applications before 2005. The solution to 4 is to keep the record that has been housed. If there are multiple housed records, the record with the maximum date is kept. The reason this is most problematic is if housed and not housed applications for the same household are kept in the data, there is potential to treat the same household as a case and a counterfactual.	Not really, just a result of messy data and processes.

Date	Decision	Description	Implications	Could this be improved for the next iteration? If so, provide details
11/08/2016	To remove transfer applications from the 2005/06 cohort. Note: This contradicts the decision made by Person A and Person B two weeks ago as it doesn't technically align with policy's original question (ROI on social housing policy process (all decisions, new applications and transfers) in 2005/06). But given further information, this decision makes the analysis much easier so we have changed the decision. It has also been run by policy (Person C) who has said removing transfer applications is fine.	The cohort will only include new applications in 2005/06, who have been housed within two years. Transfer applications are approx 10% of all applications in 2005/06	From an analytical perspective, transfer applications are not the same as new applications. They have had previous exposure to social housing, and also, regardless of their transfer application outcome (housed or not housed), they are still always housed (a case). This means, for a propensity match, there is nothing to predict and therefore you cannot create a comparison group for them to use to calculate ROI. One way around this would be to treat transfers as new applications who are housed. This would work as a rough fix for the propensity match, but could impact in the ROI calculation. If it is assumed that the return through time after being housed is not the same from year to year, such as more benefits are accrued in the first year following on from being housed, these benefits will be missed if transfers are treated as new applications. This is because they have had previous exposure to social housing so benefits won't be counted from the true date they are first housed. Transfer applications are more similar to people already housed before 2005 (who are not included in this analysis), than they are with new applications in 2005/06. This is another argument for removing them from our cohort. From a policy perspective, the question is: *what is the ROI for the social housing process in 2005/06?* This includes all decisions made for social housing in 2005/06 relating to investing (new applications and transfers). Removing transfer applications changes the question slightly to: *what is the ROI for the social housing process in 2005/06, relating to new applications*? Policy has agreed this makes more sense, as from an ROI Through time perspective this is more accurate analytically. It will also make the analytical output more straightforward, and easier to describe.	No. Transfer applications should be treated as those already in a house, that is, have had prior exposure to the intervention.

Date	Decision	Description	Implications	Could this be improved for the next iteration? If so, provide details
11/08/2016	Durations that are greater than our profile or forecast window	Durations that are greater than our profile or forecast window are possible in two scenarios: 1. Duplicate individuals in the waitlist spells (about 2% of people are in multiple households). 2. Overlapping events in the event table	To fix this, variables showing event durations will have a max value for modelling purposes (1461 days for our profile period of four days). This won't be fixed in the SQL rolled-up tables, Person A and Person B will do this as part of their data cleaning for the model as it will be faster since most of the tables have been rolled-up now. A max value limit for duration to the macro could be added, but as above the tables are almost all rolled-up, so it is not worth going back to do this. This is on the to-do list before the macro is shared, and it will be clearly stated that it doesn't support overlapping events of the same type.	
19/08/2016	To use the same tax amount for all W&S	The same tax amount will be applied to all W&S earned by members of our cohort. This decision is a time-saving one only.	The implication is the tax amount will be incorrect, but under extreme time pressure it is the best and fastest way to calculate the tax component. A rate of 13.39% of the six years was applied to everyone in the cohort. This was the average tax amount for the lowest two brackets over the six years.	Yes, definitely.
1/11/2016	Add additional enhancements to the Social Housing test case based on feedback received	There are several changes, signed off by MSD and SIU. If it is found that some of these changes are not possible then it will be noted in this log.	The feedback received will ensure our work is more robust and that the technical report is clearer to read.	There are always improvements but we will have made the bulk of the improvements suggested.
10/11/2016	Our population size is the 21,828	Given constraints for iteration 2 it was decided we would use a subset of our original population our new business rules around duplicates and attaching to the spine meant we arrived at 21,828 people. When we calculate our costs we will use all 21,828, even though some have zero costs.	It turns out 42 of the applications have zero costs. These people were in the HNZ records but not in IR or MoH records. This could be due to linkage area and would be well within the ~1.4% linkage error on the spine. Alternatively it could be a primary applicant who has deceased or a household that has emigrated.	Last time we excluded them when calculating averages. This time we will include them as it is possible that they are zero cost applications. This is in line with the feedback we received from our first version of the technical report.

Appendix J Caveats, limitations and assumptions

1. The cases and counterfactuals in this report are statistically matched based on available information on the date of application. Most of this information changes over time, so this should be considered with all analysis. The match accuracy may decrease with time, depending on how quickly the information used changes. For example, age groups are as at the time the application for housing was made – therefore it is possible to shift into next age bracket during the ROI window.
2. Where cost data was not available it was derived using the best information available at the time. See Section 2.5.1 Derived cost/return information in the IDI for details of these derivations.
3. Data from different agencies is of varying quality. Refer to SIAL and its accompanying documentation.
4. 'Household' in this analysis is defined as the group of people attached to a single HNZ application.
5. All analysis is based at the household level, as the match of cases and counterfactuals is at the household level.
6. It was assumed an applicant household is 'housed' the day after it exits the application register table. This means costs are counted from this date. However, this is not always the case – some households can wait weeks to be housed but, from a coding point of view, this assumption proved the most efficient.
7. The length of tenure in a social house, i.e. exposure time to the intervention, is not taken into consideration in this analysis.
8. ROI by year is not included as social housing has an on-going intervention cost, which makes this a methodologically difficult task.
9. Social housing occupancy outside of 2005/06 was not checked. Therefore there is the possibility households in either the case or counterfactual group were also housed before 2005 and exited the house before 2005. For the comparison group, there is also the possibility they were housed after 2006. The latter is more problematic but is not something that would have been known at the time of the social housing operation decision in 2005/06, so it was not factored into the match.
10. This evaluation measures the ATT for the process of social housing between 2005/06, not all social housing.
11. Only taxes from W&S were included in the revenue. Self-employment is not captured in these results.
12. An average tax rate of 13.39% has been applied to all W&S to calculate the tax component. This rate was derived by averaging each tax band over the time those who were in social housing were followed. This is a simplifying assumption chosen due to time constraints.
13. Findings are preliminary and subject to a number of limitations due to the short timescale of data available. It will be important for future work to develop methods for monetising the benefits of expenditure in the long-run, to produce a true lifetime ROI, and to better incorporate the findings within the social housing investment approach.
14. In addition to these initial results, the costs spent on social housing tenants compared to the counterfactual can be analysed in much more detail. For example, within agencies costs can be analysed, as well as agency costs by segments. SIU is available to support agencies with this analysis, by providing results, code and support.

15. CHPs and council housing providers were not part of this analysis.
16. Totals do not necessarily equal the sum of their components, due to confidentiality rounding.
17. No outliers were removed from the analytical output.
18. The returns calculated in this test case are fiscal only based on government administrative data contained in the IDI. This narrow focus is because the IDI is the best integrated data source available for New Zealand's population across time. It will be important to develop social, economic and, if appropriate, cultural ROI measures to complement this fiscal insight.
19. It was decided to include changes in spending on the AS as a benefit, but not to include IRRS, because the test case is assessing the effectiveness of social housing, not all social housing assistance. Furthermore, those who leave social housing can receive an AS.
20. Care is necessary using counts of CYF and corrections events. Sometimes several events can be linked to a single spell. For example, a child who has been placed in care could shift between different family members (e.g. aunt, grandparents etc), over a short period of time. This would be counted as several different CYF events when it refers to one period of care. It would be wrong to interpret the count as literally the number of interactions with CYF. However, it does indicate a greater potential of vulnerability for the child.

The variables on which the case and counterfactual groups are balanced are calculated as at the date of application for a social housing. This can have an effect on the interpretation of the results. For example, regarding age group, applicants can potentially shift in to the next age band over the course of the analysis period, (i.e. they become right censored).